Representing Mathematical Concepts Associated With Formulas Using Math Entity Cards

by

Abishai Dmello

THESIS

Presented to the Faculty of the Department of Computer Science

Golisano College of Computer and Information Sciences

Rochester Institute of Technology

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Computer Science

Rochester Institute of Technology

October 2019

Representing Mathematical Concepts Associated With Formulas Using Math Entity Cards

APPROVED BY

SUPERVISING COMMITTEE:

Dr. Richard Zanibbi, Advisor

Dr. Carlos Rivero, Reader

Dr. Matthew Fluet, Observer

Acknowledgments

This thesis has been an adventurous experience, one I am very glad I decided to take up. I would like to thank a lot of people who helped make the journey memorable.

First, I would like to express my deepest gratitude to my advisor Dr. Richard Zanibbi for being encouraging and fun to work with. From giving me the opportunity to work in the DPRL, to believing in my idea and guiding me throughout the way. Thanks to Dr. Carlos Rivero and Dr. Matthew Fluet for being on my committee and for their constructive feedback.

Thanks to Prof. Jian Wu, who suggested that we first concern ourselves with fetching definition as they are and then later focus efforts on improvising the method with the help of Machine Learning algorithms.

Thanks to Katherine Zanibbi for her review and advice on the experimental design. Prof. Anurag Agarwal for his guidance from a Mathematician point of view. Dr. C. Lee Giles (The Pennsylvania State University) and Douglas W. Oard (University of Maryland) for their feedback and support on the idea.

Special thanks to Behrooz & Gavin for the creative discussions and helping put together the entire system. Other members of the DPRL: Jennifer, Mahshad, Parag, Puneeth and Wei for the productive discussions. Thanks to Shaurya and Neisarg for their suggestions during our meetings.

Thanks to my brother Zak, family and friends who have checked and cheered me on. Forever grateful to my parents Daniel and Lydia Dmello, whose unconditional love and encouragement have shaped me into the person I am today.

Finally thanks to Smitha, my loving wife for hearing me out time and again and for being a pillar of strength throughout my entire journey.

Abstract

Representing Mathematical Concepts Associated With Formulas Using Math Entity Cards

Abishai Dmello, M.S.
Rochester Institute of Technology, 2019

Supervisor: Dr. Richard Zanibbi

We introduce Math Entity Cards, a modified version of existing Entity Cards specifically tailored for Math Information Retrieval. Math Entity Cards help connect formulas to titles and description and make the navigation between formulas and text related to formulas, seamless. These cards are populated from a new knowledge base, created by extracting and combining formulas, titles and descriptions from three different sources, Wikidata, Wiktionary & ProofWiki. We demonstrate a novel approach of using entity cards for auto-complete by integrating our cards into a Math-Aware Search Interface: MathSeer. This helps create a new ecosystem for consuming information during formula editing and search. We design and conduct a human experiment, in a math information retrieval setting and find statistical evidence for the usefulness of individual card components.

Table of Contents

Acknowledgments **iii**

Abstract **v**

List of Tables **x**

List of Figures **xii**

Chapter 1. Introduction **1**

1.1 Mathematical Concepts as Entities 3

1.2 Problem Statement & Contributions 5

1.2.1 Math Entity Card Proposed Use Case by User Search Needs 6

Chapter 2. Related Work **8**

2.1 What is an Entity? 10

2.2 What are Entity Cards? How are they created? 11

2.2.1 Entity Card Creation 14

2.3 Math Entity Cards 15

2.3.1 Information Extraction From Surrounding Text 16

2.3.2 Math Entity Linking 22

2.3.3 Mathematical Knowledge Base Creation 24

2.4 Summary 25

Chapter 3. Math Entity Card Design **26**

3.1 Formula Description Card Designs 27

3.2 Math Entity Card: Additional Usage Section 29

3.3 Math Entity Cards for Symbols 31

3.4 Alternate Descriptions for a Concept or Formula 33

3.5 Concept Titles & Aliases 36

3.6 Math Entity Cards in a Math Aware Search Interface 37
3.7 Summary 40

Chapter 4. Math Entity Card Creation 41
4.1 Math Entity Card Creation 41
4.1.1 Extracting Formulas & Titles From Structured Data Sources 43
4.1.2 Adding Mathematical Concept Descriptions Using Wikipedia 45
4.1.2.1 Extracting Symbol Content 46
4.1.3 Extracting Formula, Title & Defintions From Wiktionary 46
4.1.3.1 Selecting A Single Math Expression From Wiktionary Definition 49
4.1.4 Extracting Formal Mathematical Definitions From ProofWiki 49
4.1.4.1 Selecting A Single Math Expression From ProofWiki Definition 52
4.2 Synthesizing the Data 54
4.2.1 Math Entity Card Prototype & API for Auto-complete . 55

Chapter 5. Human Experiment 60
5.1 Experiment Design 61
5.1.1 Mathematical Entity Selection 64
5.2 Participants 64
5.3 Variables & Confounds 65
5.4 Procedure 66
5.5 Trials 67
5.6 Post-Study Questionnaire 69
5.7 Summary 70

Chapter 6. Results 71
6.1 Demographics 71
6.2 Experiment 75
6.2.1 Usefulness of Card Components 77
6.2.2 Understanding of Content 78
6.2.3 Participant Comments 81
6.2.4 Secondary Results 83

6.2.4.1 Between Familiar & Less Familiar Concepts . . 84
6.2.4.2 Between Symbols, Small and Large Formulas . . 84
6.3 Statistical Testing 84
6.3.1 Usefulness score 85
6.3.2 Response Times 87
6.4 Post Study Questionnaire 88
6.4.1 Participant Comments 90
6.5 Discussion 92
6.6 Summary 95

Chapter 7. Conclusion and Future Work 97
7.1 Contributions 98
7.2 Future Work 99
7.2.1 Data Quality 100
7.2.2 Additional Benefits 101
7.2.2.1 Use of Computer Algebra System 101
7.2.2.2 Tutorial Links & Related Work 101

Appendices 104

Appendix A. Recruiting Email 105

Appendix B. Recruiting Poster 107

Appendix C. Pre-Written Script 108

Appendix D. Card Types in Human Experiment 110
D.1 Symbols 110
D.2 Small Formulas 122
D.3 Large Formulas 131

Appendix E. Mathematical Concepts as per familiarity in Human Experiment 139
E.1 Symbols 139
E.2 Small Formulas 140
E.3 Large Formulas 140

Appendix F. Secondary Results **141**

Bibliography **145**

List of Tables

2.1 Examples of candidate triples from the selection process . . . 16

2.2 Examples of extracted patterns from candidates after the surface level text matching process 18

2.3 Subset of Features used for Machine learning 19

2.4 E.g. of Sentence Patterns 21

2.5 Textual Span Definitions 22

4.1 Distribution of Formulas per Concept 45

4.2 Match percentage between title and strong tag contents in Wiktionary 48

4.3 Number of Math Per Description Wiktionary Note : Titles are not exclusive, some titles have multiple descriptions 48

4.4 Multiple Descriptions for the Binomial Coefficent from different data sources. 54

5.1 Descriptive Statistics of formulas sizes (without single symbols) in the Wikidata Data set 63

5.2 Counterbalanced order used to present card types to participants (P) for corresponding mathematical entities (E) 65

5.3 Questions from Section 1 of Post Study Questionnaire. 69

5.4 Questions from Section 2 of Post Study Questionnaire. 70

6.1 Pairwise Wilcoxon Signed Rank Test with Bonferroni Correction for usefulness scores 86

6.2 Observations : Pairwise Wilcoxon Signed Rank Test with Bonferroni Correction 86

6.3 Pairwise T-Test with Bonferroni Correction for response times 87

6.4 Observations : Pairwise Wilcoxon Signed Rank Test with Bonferroni Correction in Response times 88

E.1 Familiar and Less Familiar, Symbols used for Human Experiment. *Indicates Practice Trials 139

E.2 Familiar and Less Familiar, Small Formulas used for Human Experiment. *Indicates Practice Trials 140

E.3 Familiar and Less Familiar, Large Formulas used for Human Experiment. *Indicates Practice Trials 140

List of Figures

1.1 Example of Search Results as presented by Approach0 - a math aware search engine 2

1.2 Examples of an entity card on different search engines for a common query : ‘Pythagorean Theorem’ 4

1.3 Math entity card template 5

2.1 Example of entity card displayed on the Google SERP for different queries 11

3.1 Similarities & differences in layout between common entity card as described by Balog K. [2] and proposed math entity card. . 27

3.2 Examples of math entity cards with title and formula only. Wikipedia indicates the source URL. 27

3.3 Examples of math entity cards with title-formula and descriptions/definitions 29

3.4 Examples of math entity cards with title-formula-description and a Usage section 30

3.5 $\sin\theta$ card with related functions/operations as usage 30

3.6 Math entity cards for mathematical symbols 31

3.7 Faceted Search for Symbol Cards 33

3.8 Binomial Coefficient with multiple formulas and and multiple Descriptions 34

3.9 Stacked Cards with Pop Up 35

3.10 Normal Distribution Card with Aliases 36

3.11 LaTeX formula auto-complete as present in Wolfram Alpha . . 38

3.12 Math entity Cards as auto-complete 40

4.1 A section of Pythagorean Theorem from Wikipedia, highlighted are multiple valid definitions. 42

4.2 ProofWiki page for Binomial Coefficient with math highlighted in Red Boxes 53

4.3 Prototype of Math Entity Cards 56
4.4 Prototype of math entity cards with Carousel for rotating descriptions. 57
4.5 A subset of formulas, along with other fields as stored in the database 58

5.1 Different combinations of card components (Title, Formula, Description) forming different card types. 60
5.2 Card for Addition containing title and description 68

6.1 Age & education of participants 71
6.2 Bar plot of math courses taken and frequency of looking up mathematical information as reported by participants 73
6.3 Frequency with which participants need to express mathematical notation 74
6.4 Box plot of overall and indiviual time taken for section 1 of each trial 75
6.5 Overall Usefulness Scores per Card Type 77
6.6 Card-types contribution to understanding, for familiar concepts 78
6.7 Card-types contribution to understanding, for less familiar concepts 80
6.8 Comment Distribution across Groups 81
6.9 Comment Distribution across Card Types 83
6.10 Importance of title, formula and description 89
6.11 Usefulness of related concepts, tutorials and formal description 90

7.1 A complete math entity card design 102

D.1 Card Types for Congruence 110
D.2 Card Types for Inequality 111
D.3 Card Types for Line Integral 111
D.4 Card Types for Complex Conjugate 112
D.5 Card Types for Cross Product 113
D.6 Card Types for Aleph Number 114
D.7 Card Types for Converse Implication 114
D.8 Card Types for Projective Space 115

D.9 Card Types for Compact Embedding 116
D.10 Card Types for Partial Derivative 116
D.11 Card Types for Plus-Minus 117
D.12 Card Types for Left Open Interval 117
D.13 Card Types for Entailment 118
D.14 Card Types for Beth Number 119
D.15 Card Types for Wreath Product 120
D.16 Card Types for Covering Relation 121
D.17 Card Types for Adsorption 122
D.18 Card Types for Autonomous Consumption 122
D.19 Card Types for Rotating Unbalance 123
D.20 Card Types for Classification Of Electromagnetic Fields 123
D.21 Card Types for Reality Structure 124
D.22 Card Types for Magnetic Energy 124
D.23 Card Types for Mired 125
D.24 Card Types for Allan Variance 125
D.25 Card Types for Angular Velocity 126
D.26 Card Types for Equianharmonic 126
D.27 Card Types for Huge Cardinal 127
D.28 Card Types for Ratio Test 128
D.29 Card Types for Divisor 129
D.30 Card Types for Solenoid 129
D.31 Card Types for Conformational Isomerism 130
D.32 Card Types for Ch´zy Formula 130
D.33 Card Types for Rayleigh Distribution 131
D.34 Card Types for Bernoulli's Inequality 131
D.35 Card Types for Lower Hybrid Oscillation 132
D.36 Card Types for Sine 132
D.37 Card Types for Phase Retrieval 133
D.38 Card Types for Electrostatic Force Microscope 133
D.39 Card Types for Integral Equation 134
D.40 Card Types for Dew Point 134

D.41 Card Types for Oscillatory Integral 135
D.42 Card Types for Gumbel Distribution 135
D.43 Card Types for Klecka's Tau 136
D.44 Card Types for Epimorphism 136
D.45 Card Types for Optical Transfer Function 137
D.46 Card Types for Lee Distance 137
D.47 Card Types for Parallelogram Law 138
D.48 Card Types for Antenna Gain To Noise Temperature 138

F.1 Distribution of Usefulness Scores across total number of cards for Familiar Concept 141
F.2 Distribution of Usefulness Scores across total number of cards for Less Familiar Concept 142
F.3 Average Usefulness Scores per Card Type for Familiar Concepts 142
F.4 Average Usefulness Scores per Card Type for Less Familiar Concepts 143
F.5 Distribution of Usefulness Scores across total number of cards for Symbols 143
F.6 Distribution of Usefulness Scores across total number of cards for Small Formulas 144
F.7 Distribution of Usefulness Scores across total number of cards for Large Formulas 144

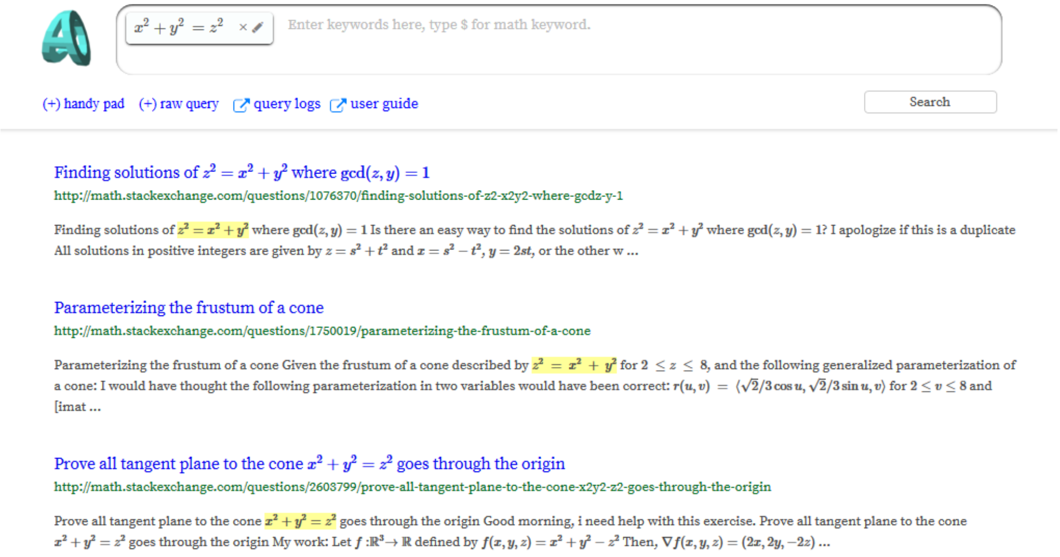
Chapter 1

Introduction

Navigating currently between formulas of mathematical concepts and their associated names or descriptions is a rather long and sometimes tedious process. Math Entity Cards are designed to help make this transition from formula to concept as well from concept to formula, simple and straightforward.

Mathematical formulas are a part of the abstraction process, they have both syntax and semantics and are widely used to convey some information, just like text. However, existing text search engines are not built to support mathematical formulas. They either treat formulas as LaTeX strings or assume meaning based on surrounding text or ignore the math altogether. Thus when searching for mathematical concepts, this leads to longer search sessions, increase in the number of query reformulations and an overall decrease in the search experience.

Contrary to text search engines that focus primarily on text, math search engines revolve mainly around formulas as an input while also supporting text based search. Existing math information retrieval (MIR) systems such as Approach0 [37], Tangent [26] and WikiMirs [11] display their results in a manner similar to text based retrieval systems (Figure 1.1), by listing URLs



Finding solutions of $z^2 = x^2 + y^2$ where $\gcd(z, y) = 1$

http://math.stackexchange.com/questions/1076370/finding-solutions-of-z2-x2y2-where-gcdz-y-1

Finding solutions of $z^2 = x^2 + y^2$ where $\gcd(z, y) = 1$ Is there an easy way to find the solutions of $z^2 = x^2 + y^2$ where $\gcd(z, y) = 1$? I apologize if this is a duplicate All solutions in positive integers are given by $z = s^2 + t^2$ and $x = s^2 - t^2$, $y = 2st$, or the other w ...

Parameterizing the frustum of a cone

http://math.stackexchange.com/questions/1750019/parameterizing-the-frustum-of-a-cone

Parameterizing the frustum of a cone Given the frustum of a cone described by $z^2 = x^2 + y^2$ for $2 \leq z \leq 8$, and the following generalized parameterization of a cone: I would have thought the following parameterization in two variables would have been correct: $r(u, v) = \langle \sqrt{2}/3 \cos u, \sqrt{2}/3 \sin u, v \rangle$ for $2 \leq v \leq 8$ and [imat ...

Prove all tangent plane to the cone $x^2 + y^2 = z^2$ goes through the origin

http://math.stackexchange.com/questions/2603799/prove-all-tangent-plane-to-the-cone-x2y2-z2-goes-through-the-origin

Prove all tangent plane to the cone $x^2 + y^2 = z^2$ goes through the origin Good morning, i need help with this exercise. Prove all tangent plane to the cone $x^2 + y^2 = z^2$ goes through the origin My work: Let $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(x, y, z) = x^2 + y^2 - z^2$ Then, $\nabla f(x, y, z) = (2x, 2y, -2z)$ ...

Figure 1.1: Example of Search Results as presented by Approach0 - a math aware search engine

and a small snippet of content that has the matched portion of the query highlighted. While this is beneficial for regular or exploratory search, it does not help look up factual information. That is to say if a user has entered a formula that defines or is related to a mathematical concept or theorem, only highlighting relevant/partial matches might miss out on addressing the search intent, which is probably to know more about the concept or formula.

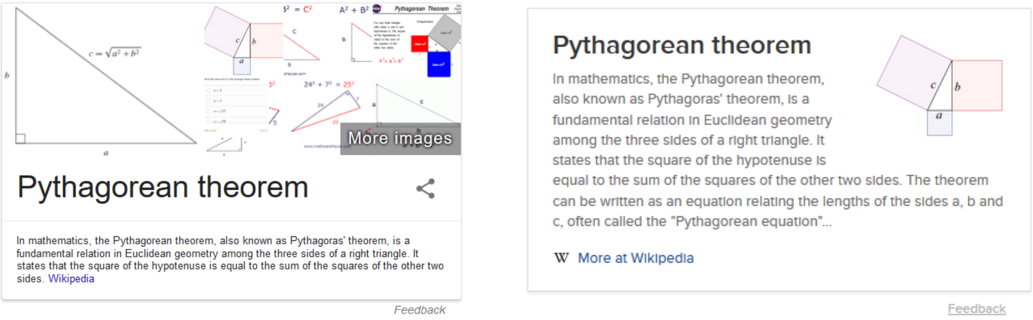
A few years ago text-based search engines faced a similar issue, but have evolved from simple matching of text keywords, to now analyzing queries to better understand and respond to a user's information need. One way they do so, is by supplementing the Search Engine Result Page (SERP) with additional results based on an educated guess of what a user is looking for. For example, if we query the phrase “Albert Einstein” on any commercial text-search engine, the results describe the famous theoretical physicist, by providing information on who he was and what he accomplished, rather than just sources of text where the phrase “Albert Einstein” occurs. Balog K. defines this approach of returning information about entities (real world uniquely identifiable objects) as Entity Oriented Search [2]. This behavior of search engines hence reflects an understanding of query terms where information collected about real world entities is fetched based on relationships between what is being asked, and what is already known about the entity.

1.1 Mathematical Concepts as Entities

There are certain mathematical equations and concepts that are more familiar to users than others, e.g. ‘Pythagorean Theorem’ which is usually represented by the equation

$$a^2 + b^2 = c^2 \tag{1.1}$$

If we search for the text phrase ‘Pythagorean Theorem’ in a commercial search engine, along with the regular results we are provided with a small info-box also called an Entity Card. Figure 1.2a and 1.2b, each provide an example of the entity card for two common text search engines Google and DuckDuckGo. As we see both cards have the same description for the Pythagorean theorem, a common image and a link to the common source of extraction, i.e. Wikipeida. This extraction of entity cards for text search engines naturally follows a text-based or text first approach, of matching keywords to pages and



Pythagorean theorem

In mathematics, the Pythagorean theorem, also known as Pythagoras' theorem, is a fundamental relation in Euclidean geometry among the three sides of a right triangle. It states that the square of the hypotenuse is equal to the sum of the squares of the other two sides. Wikipedia

Pythagorean theorem

In mathematics, the Pythagorean theorem, also known as Pythagoras' theorem, is a fundamental relation in Euclidean geometry among the three sides of a right triangle. It states that the square of the hypotenuse is equal to the sum of the squares of the other two sides. The theorem can be written as an equation relating the lengths of the sides a, b and c, often called the "Pythagorean equation"...

W More at Wikipedia

(a) Google

(b) DuckDuckGo

Figure 1.2: Examples of an entity card on different search engines for a common query : ‘Pythagorean Theorem’

extracting general descriptions from the page. However from a mathematical search perspective the card does not have either a **formula**, **description of the formula and its variables**, or **applications of the concept**, which we believe is crucial in addressing a user’s math informational need. With a few more clicks and effort to filter through some more information a user would possibly find the formula, its description and the corresponding applications.

Math-aware search engines on the other hand, revolve mainly around formulas as inputs, we hence describe a process of using formulas as starting point to fetch names (titles) and descriptions of concepts to which these formulas act as attributes. We describe this to be a formula first approach by working our way from formulas to concepts instead of concepts to formulas, the latter which as seen before although possible in existing search engines is time consuming.

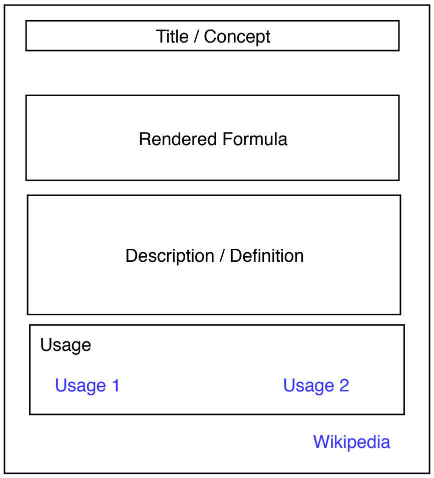


Figure 1.3: Math entity card template

1.2 Problem Statement & Contributions

This thesis aims to explore the following research questions:

1. If mathematical concepts are entities, can formulas be associated with them? If yes, can we use entity cards to navigate between formulas and concepts?

2. Would providing more mathematical information during search be beneficial to users?

In order to address the research questions, the following contributions are made as part of this thesis:

1. An alternate design of entity cards (Figure 1.3), specifically meant to address various types of mathematical search needs, that current entity cards for text-based mathematical search do not address.

2. Populating individual components (title, formula and description) of these cards by compiling data from existing structured and semi-structured data-sources.

3. A human experiment to study the usefulness of individual card components while searching for mathematical content from both a text query and a LaTeX query input.

4. Creation of an index on both titles and formulas, that can be queried via an API, and demonstrating an alternate use of these cards as a form of auto-complete.

1.2.1 Math Entity Card Proposed Use Case by User Search Needs

Zhao et al. [36] were the first to categorize math user's needs into **informational needs**: searching for a name/alias, definition, derivation, explanation, application etc. and **resource needs**: searching for paper, tutorial, slides etc. However based on a taxonomy of web search goals as created by Broder [4] there exists a third web-search need that is relevant to math search as well, **a navigational need**. The purpose of a navigational need is to re-find the exact page/document containing the formula, that was previously encountered.

For a beginner looking for the concept associated to an unknown formula or for an expert looking for a precise technical description of either a concept or a formula, Math entity cards can help address this **informational need**. For an expert looking to understand other related concepts connected to a concept of interest or a beginner looking for a tutorial of the existing concept, Math entity cards could help address this **resource need**. Math entity cards in general help provide a two way access of navigating to either the concept from the formula or the formula from the concept, thus addressing a **navigational need**.

We first introduce the existing work on entity cards and their studied effects in text search engines followed by the work done in extracting descriptions for mathematical formulas. We then provide our modifications to the existing designs of entity cards to create math entity cards. Rather than extracting title, formula and descriptions triples from sentences as done in the previous work, chapter 4 discusses about methods in which these card components can be populated by compiling data from existing sources. It also describes how by creating a dual index on both formulas and titles, these cards are used as auto-complete in MathSeer. Chapter 5 describes the human experiment carried out to observe the usefulness of individual card components (title, formula & descriptions), in isolation without any search interface. Chapter 6 describes our results and observations from the human experiment. Finally we discuss future opportunities and areas to improve upon.

Chapter 2

Related Work

Zhao et al. [36] propose the notion of ‘Keyword-to-Expression Linking’ i.e. the resolution of expressions to terminology (e.g $a^2+b^2 = c^2$ to *Pythagorean theorem*) as a means to bridge the gap, between making expression searching and relevance ranking relevant to users while maintaining the usability of keyword searches in text-search engines. Sapa et al. [30] in their user study on information seeking behaviour of mathematicians, scientists and students, observe that students search more often for reference works (encyclopedias and dictionaries etc.) and more often use, search engines designed to find specific objects (e.g. graphics, audio files, multimedia objects). Although this could be a result of the need of learning or homework activities, they do classify it as both an informational and resource need. They also found a majority of both students and scientist starting their math information search from Google, a text based search engine.

Mansouri et al. [23] were the first to characterize searches for mathematical concepts from search engine query logs. Apart from longer search sessions they found that math queries are considerably longer on average than typical web queries and have long runs of cut-and-paste text. They also found amongst the requested content, tutorials in any form (text, slides, videos or any combination) were the most frequently requested content type followed by PDF and video. Based on the frequency of question type keywords in math queries they found ‘What’ followed by words such as Formula (60%), Equation (11%) and Used for (9%) to be occurring in 69.5% of queries. This by Zhao’s definition demonstrates that a considerable amount of math based information needs are informational in that, the search is mainly for data that can be considered as facts related to a mathematical concept.

Long length of math queries, extensive query refinement and longer search sessions also results in lower satisfaction levels as predicted by Mansouri et al. [23]. This in some way could be attributed towards search engines not being able to interpret/understand what exactly is being asked. Text based search engines do not deal with mathematical expressions as well as they deal with text queries, the reason for this is firstly, the input to these engines are purely text based, which means users would have to resort to either entering LaTeX for mathematical expressions or using some set of keywords for mathematical terms. This like Zhao et al. [36] and Wangari et al. [33] studied, results in an expression gap between users and search systems. Users spend more time on creating a query and reformulating it in a manner that the search engines understands and can then return results that are meaningful to the user.

Search however is only a part of the process, when an information need arises, it is not the end. Text-based search engines are constantly working on innovative ways to understand user queries and present information in ways that are more readily consumable. This chapter describes some of the ways text-based search engines are doing so and draws connections to previous work in math information retrieval which when combined could be applied to improve how users search and consume math information.

2.1 What is an Entity?

Balog K. in his book on Entity-Oriented Search [2] defines an entity to be a uniquely identifiable object or thing, that can be characterized by its name(s), type(s), attributes, and relationships to other entities. The author goes onto further classify entities into

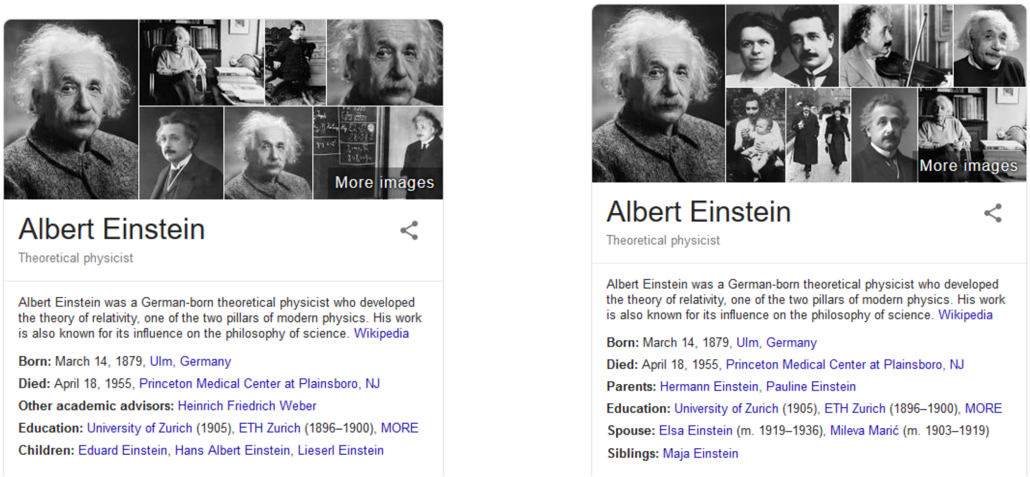
- Named Entities: which are entities that can be mapped to a real world object e.g., Albert Einstein or Golden Gate Bridge.
- Concepts: Abstract objects that map to mathematical, philosophical, physical, psychological social concepts or sometimes even natural phenomena, e.g., Triangle, Conscience or Earthquake.

The author also mentions that from previous studies on query logs, about 40-70% of queries issued to general text search engines either have an entity mentioned or target some specific entities. Mansouri et al. [23] had conducted their study by identifying mathematical entities represented as text keywords in query logs. They found approximately 400,000 queries out of 27 million records that contained at least one distinctive mathematical term (e.g. ‘Taylor

Series'). This supports the idea of "Entity Oriented Search" as coined by Balog, for math Information Retrieval as well.

2.2 What are Entity Cards? How are they created?

Search engines such as Bing, DuckDuckGo, and Google have started responding to queries containing identifiable entities such as "Einstein Education" or "Albert Einstein Family" with Entity Cards also known as summary cards (Figure 2.1a & Figure 2.1b). The entity cards appear on top right hand side of the Search Engine Result Page (SERP) so as to supplement the other search results (10-blue links) for a query.



Albert Einstein

Theoretical physicist

Albert Einstein was a German-born theoretical physicist who developed the theory of relativity, one of the two pillars of modern physics. His work is also known for its influence on the philosophy of science. Wikipedia

Born: March 14, 1879, Ulm, Germany

Died: April 18, 1955, Princeton Medical Center at Plainsboro, NJ

Other academic advisors: Heinrich Friedrich Weber

Education: University of Zurich (1905), ETH Zurich (1896–1900), MORE

Children: Eduard Einstein, Hans Albert Einstein, Lieserl Einstein

(a) Query : 'Einstein Education'

Albert Einstein

Theoretical physicist

Albert Einstein was a German-born theoretical physicist who developed the theory of relativity, one of the two pillars of modern physics. His work is also known for its influence on the philosophy of science. Wikipedia

Born: March 14, 1879, Ulm, Germany

Died: April 18, 1955, Princeton Medical Center at Plainsboro, NJ

Parents: Hermann Einstein, Pauline Einstein

Education: University of Zurich (1905), ETH Zurich (1896–1900), MORE

Spouse: Elsa Einstein (m. 1919–1936), Mileva Marić (m. 1903–1919)

Siblings: Maja Einstein

(b) Query :'Albert Einstein Family'

Figure 2.1: Example of entity card displayed on the Google SERP for different queries

Entity cards are a concise, independent (from the SERP by appearing on the right hand side of the search results), collection of information including a title/name, possibly an image and a summary: a set of facts from an underlying knowledge base, all that describe the entity [9]. In (Figure 2.1a & Figure 2.1b) we notice both the queries have ‘Einstein’ in common, which is considered to be the common entity.

Studies by Bota et al. [3] have attempted to answer questions such as

- How does the card presence and content influence users’ search behaviour and perceived workload?
- Do card properties, such as card coherence (whether card contents are coherent and all focus on the same topic of a user’s query) and vertical diversity (whether cards contain visually salient blocks of elements, such as Images), have an effect on search behaviour and workload?

By conducting a large scale crowd study they have been able to measure and analyze the following:

- Card Interactions, which refers to how users engage with entity cards containing both on topic and off topic content.
- Web Interactions, which focuses on searchers engagement with non-paid/non-advertised (organic) web results displayed on the SERP.
- Workload, which focuses on the perceived task load as measured by a post study questionnaire.

Their study find differences in user's interaction with entity cards and search results due to on-topic and off-topic card contents. They found searchers spend less time interacting with organic web results when the entity card is off topic compared to being on-topic or even absent. Their studies verify a logical assumption of searchers issuing less modified queries when the entity card is on-topic as compared to off-topic. With respect to the workload, their study finds on-topic entity cards do not affect perceived workload as compared to absence of entity cards, however off topic entity cards could generate more workload because of the additional information users need to examine.

Entity cards are present not just for regular queries but also for queries containing health related conditions. Consumer Health Search (CHS) is described to be a challenging domain with challenges such as vocabulary mismatch, and lack of domain expertise which affect both query formulation and result interpretation. Recent user studies in domain specific entity cards by Jimmy et al. [14], have found Health Cards being able to help less knowledgeable users search and diagnose health conditions as effectively as more knowledgeable users. They conclude that Health Cards are best suited for well-defined health search tasks (e.g.Factual Scenarios) rather than exploratory tasks. In a follow up study Jimmy et al. [15] investigate the effectiveness of Health cards to assist in decision making in CHS, where in they propose a novel multi-card interface. A multi-card interface shows multiple cards all stack adjacently to allow users to perform comparison based diagnosis (differential diagnosis). They conclude that the multi-card interface helps users to make health decisions such as correct diagnosis and predicting the urgency of treatment with significantly lesser effort than a single card. The challenges faced by CHS however is analogous to math information retrieval and many other domain specific information retrieval scenarios where in users might know the exact term to query and hence approximate the query by self describing the situation. This more often than not, results in users modifying the query and repeating the search to narrow down results. To help with CHS, there is also the development of tool or info-tip with entity card like functionality by Lopes et al. [21] to Assist Health Consumers while searching for the web by providing Medical Annotations. The tool annotates medical concepts present on a web page and allows access to information such as concept definition, related concepts and links to external references for these annotated concepts.

2.2.1 Entity Card Creation

Text-based search engines such as Google and Bing make use of their own proprietary knowledge bases/graphs to generate entity cards. They do so by fetching the name/title, an image, a description or summary and a set of facts from this knowledge base, all that describe the entity [12].

In Figure 2.1a and Figure 2.1b the information on the card changes, with changes in the query, although both queries have the same entity i.e. ‘Einstein’ each entity card differs a bit in content, query for ‘Albert Einstein Family’ responds with a card containing information about his parents, spouse and children which are not present for the query ‘Einstein Education.’ This is an example of dynamic summarization where in the contents of the card are query-dependent. Studies by Hasibi et al. [9] were the first to explore the concept of dynamic summarization for entity cards. They define dynamic summarization as a two step process comprising of fact ranking and summary generation. The fact ranking step includes ranking of facts according to importance and/or relevance to terms in the query. The second step is the rendering of these facts on the entity card. Their studies find users preferring dynamic summaries, those that are query-dependent over static summaries that are query-agnostic.

2.3 Math Entity Cards

Seeing the positive effect entity cards have on text information retrieval, we assume they would carry forward to math information retrieval and hence propose the creation of math entity cards. To the best of our knowledge, this is the first work that introduces and describes the design, creation and studies the effects of these cards in math information retrieval. As we shall see there has been prior work addressing challenges in each area of card creation such Information Extraction (Title, Description/Definition), Entity Linking and Knowledge Base creation for mathematics in isolation. But the concept of using creating and using a math entity card for math information retrieval is new. We suspect this mainly since Entity Cards as a concept for text search engines themselves are a fairly recent idea and also primarily because formulas are not fully supported in standard text-based search engines.

2.3.1 Information Extraction From Surrounding Text

Quoc et al. [28] initiated work around extracting co-reference relations between formulas and the surrounding text in Wikipedia. They do so by finding textual overlaps between formulas converted to text and text descriptions around formulas. They call this approach as surface level text matching and represent it by Equation 2.1. Their work describes the extraction of a Concept, Description and Formula (CDF) triple, in which a concept is defined to be a name or a title of a formula. Their extraction process creates a candidate concept for any noun phrase in the title, section headings or text written in bold or italic in Wikipedia articles. The selection of descriptions is based on the noun phrases (NP) that occur after variations of the verb 'to be'. Examples of the candidate pairs are shown in Table 2.1.

Table 2.1: Examples of candidate triples from the selection process

Concept	**Description**	**Formula**
the sine of an angle	the ratio of the length of the opposite side to the length of the hypotenuse	sin A $= \frac{opposite}{hypotenuse} = \frac{a}{h}$
a quadratic equation	a polynomial equation of the second degree	$ax^2 + bx + c = 0$

Their work starts out by considering only those CDF triple's that lie in the same paragraph. After the generation of candidate CDF triples, surface level text matching is used to classify each candidate as true or not based on a similarity score given by Equation 2.1. Surface level text matching can be defined as a ratio of overlap between text keywords as follows

$$sim(F, C, D) = \frac{|T_F \cap T_C|}{min\{|T_C|, |T_F|\}} + \frac{|T_F \cap T_D|}{min\{|T_D|, |T_F|\}} \tag{2.1}$$

where T_F, T_C and T_D are sets of words extracted from Formula(F), Concept(C) and Description(D) respectively. The common math operators are converted to text, e.g. '+' is converted to 'plus' and '\frac' is converted to 'divide', this implies

- Math formulas are converted to a textual representation, which may cause some loss in the structural and syntactical information they carry.
- The method is not applicable to less common operators, variables and other identifiers.

Candidates are then classified as 'True' if they meet a sim(F, CD) score no larger than 1/3. Candidates that are not classified as true, are then re-examined in a second pass by using patterns generated from the Candidates that are classified as true after the surface level matching step. Table 2.2 shows examples of the extracted patterns. CONC, DESC and FORM are placeholders for Concept, Description and Formula respectively. The classified candidates are finally evaluated manually. Their best system had an accuracy of 68.33% out of 138,285 CDF candidates after manual evaluation.

Table 2.2: Examples of extracted patterns from candidates after the surface level text matching process

Pattern
CONC is DESC: FORM
CONC is DESC. In our case FORM
CONC is DESC. So, ..., FORM
CONC FORM

Yokoi et al. [34] improve upon this work by first manually constructing a reference data-set of 100 Japanese Scientific papers. With the help of pattern matching and machine learning methods they demonstrate the challenges and feasibility of fetching variable names and function definitions from surrounding natural language descriptions. Their work focuses mainly on connecting elements of mathematical expressions with their names, definitions and explanations, which they refer as mathematical mentions. For example given a sentence, “We defined the precision(P) as follows $P = \frac{W}{W+Y}$ where W is the number of extracted correct-labeled pairs and Y is that of extracted fault-labeled pairs.” The extraction process should result in: P - the precision, W - the number of extracted correct-labeled pairs and Y that of extracted fault-labeled pairs. The task is then defined to be automatically identifying such connections and validated them against the hand annotated data-set. Since this was the first work on linking formulas to descriptions, only compound nouns (combination of two independent words that has its own meaning individually) in the same sentence was considered as possible candidates for mathematical mentions. Their basic approach also presupposes that the mathematical mentions co-occur with the target mathematical expression within the **same sentence**. They also evaluate an SVM-based binary classification approach, using a set of eight manually identified patterns. Apart form the eight pattern features they make use of other linguistic cues to help in the classification. Table 2.3 shows a subset of the features used for the SVM based approach.

Table 2.3: Subset of Features used for Machine learning

Features	**Explanations**
Another mathematical expression, comma, or opening or closing brackets	Test existence of another mathematical expression, comma between the target noun and the mathematical expression.
Order	Test whether the target noun lies anterior to the mathematical expression or not.
Composition	If the target noun is a compound noun

Every feature has a binary value of whether or not the feature is present for a sample. On further analyses of their data-set we discovered a problem of class imbalance problem where in there are 3,867 positive samples and 53,153 negative samples in training and 1,193 and 16,219 negative instances; unfortunately they do not mention how they handle this situation. They propose a novel approach for an evaluation criteria: soft and strict matching. Soft matching, considers the classified result to be true if they partially match the human annotated ones. Strict matching, as the name suggest considers the classified result to be true only if they **exactly** agree with human annotated ones. Their overall F-1 score on the test data-set is 89.20 for Soft Matching vs 84.25 for Strict Matching which considering an initial approach looks very promising, however if we consider the initially pointed out limitations of a single compound noun and an imbalanced data-set we quickly realize that the practical applications of this method are low. To overcome the first challenge Kristianto et al. [20] propose a design guideline for annotating scientific papers for mathematical formula Search. They assume a single mathematical formula can have multiple descriptions. Each description could be of two types short description that specifies the type or category of the formula e.g $\log(x)$ is a function and long description $\log(x)$ is a function that computes the natural logarithm of the value x. Kristianto et al. [8] carry forward the same work for the extraction of textual descriptions from scientific papers. They describe three different approaches for extracting the definitions of mathematical expressions under the assumption that definitions are usually noun phrases.

- Nearest Neighbor.
- Pattern Matching.
- Machine Learning.

The nearest neighbor method is the baseline method and works under the assumption that the textual definition is a combination of adjectives and nouns that occur before a mathematical expression. They make use of a part of speech tagger to obtain the annotation of words (classification of words as adjectives, nouns and verbs) surrounding the expression. The pattern matching

Table 2.4: E.g. of Sentence Patterns

No.	Sentence Pattern
1.	... denoted (as | by) MATH DEF
2.	(let | set) MATH (denote | denotes | be) DEF
3.	MATH (is | are) DEF

approach tries to capture the sentence patterns (as a set of rules) in which definitions are usually mentioned in Scientific papers. Table 2.4 provides examples of the sentence patterns used in the pattern matching method. In Table 2.4, MATH and DEF symbols denote the target mathematical expression, its definition, and other mathematical expressions, respectively. The machine learning approach uses all the patterns from the pattern matching step along with some other features such as location, unigram, bigram and trigram scores etc. For the strict matching criteria they were able to achieve a precision of 73.60, recall of 30.09 and an F-score of 42.46, and for the soft matching criteria they were able to obtain a precision of 80.08, recall of 40.30 and an F-score of 53.29, while impressive their data set consists of only 14 scientific papers and hence might not have the coverage needed to support math information retrieval at a large scale.

Kristianto et al. [19] improve on their previous description extraction methods of mathematical expressions and assess the coverage of several types of textual span: fixed context window, apposition, minimal noun phrase and all noun phrases. Table 2.5 gives the explanation of each individual textual span.

Table 2.5: Textual Span Definitions

Textual Span	**Explanations**
Fixed Context Window	Ten words before and after the target expression
Apposition	A preceding noun phrase that has the same referent (apposition) relation with the target math expression
Minimal Noun Phrase	The first compound noun phrase from a complex noun phrases that contains prepositions, adverbs or other noun phrases.
All Noun phrase	All noun phrases in the target sentence.

Similar to their previous work their evaluation included two methods soft and strict matching of definitions. Where in a candidate would pass the strict matching evaluation if its position, in terms of start index and length is the same as the gold standard. And a candidate would pass soft matching evaluation if its position contains, is contained in or overlaps with the position of the gold standard description for the same expression. Their evaluation in terms of both strict and soft matching of definitions helps conclude "apposition" gives the highest F1-score, but "minimal noun phrase" and "all noun phrase" produces the highest recall. They also point out why their previous methods [25, 20, 8] work only in particular cases e.g. Expecting an expression to have all its defining terms within a specified context window.

2.3.2 Math Entity Linking

Entity linking can be described as mapping entities in unstructured free text to known entities in a knowledge base. A variation of entity linking is wikification, which identifies an entity and locates its corresponding Wikipedia article. Linking Mathematical Expressions to Wikipedia was first explored by

Kristianto et al. [17]. They formalize the idea as "Given a document d containing a set of math mentions (math expressions/formulas) $M = \{m_1, .., m_n\}$ assign each math mention m_i a Wikipedia article t_i." The method used by Giovanni et al. [17] is not purely formula/expression based, and makes use of the surrounding text as part of two enrichment steps that are performed. The enrichment steps are as follows:

- Math Enrichment
- Text Enrichment

The math enrichment step is similar to a query expansion technique where the entire math expression is split into multiple sub-expressions based on the top-level (in)equality. This is done to help increase the percentage of partial match in case there is no exact match of the query. The output of this step is a set, that includes the original math expression along with sub expressions from the split. The text enrichment step creates a concatenation of noun phrases that contain the math expression or a sub-expression along with extracted textual description of the formula, from the same input document d, based on approaches used in their earlier work [19]. After the enrichment step a new query q_i is created which contains both math and text and this is used to identify which Wikipedia article the math mention should link to.

2.3.3 Mathematical Knowledge Base Creation

Math entity cards are expected to function in a similar manner as entity linking where isolated formulas will be matched to entries in a knowledge base to fetch known factual information regarding the formula. This subsection describes work focussed at developing mathematical knowledge bases.

With the rise of XML based languages such as MathML [1], OpenMath [5] and OMDoc [16], all with a focus of supporting exchange of mathematical information over the web, there has been prior attempts to create knowledge bases that serve as a repository mathematical information although not mainly for information retrieval, but for automated theorem proving and finding proven mathematical properties [7]. There has also been attempts to translate information between different libraries [12] with a goal to make the information more machine readable.

Today's machine readable data in knowledge bases [27] are stored in an inter-operable format such as Resource Description Framework (RDF) also known as Linked Open Data. RDF use statements to define and capture relationships between objects. The statements are stored as triples of the form subject-predicate-object. Nevzorova et al. [24] experimented with similar methods of proximity based matching of mathematical variables with noun phrases described earlier, to try and get math data to Linked Open Data. They were able to get 68% accuracy in picking formulas and their defining terms on 300 papers. This is a relatively small sample to use as a knowledge base for math entity cards.

2.4 Summary

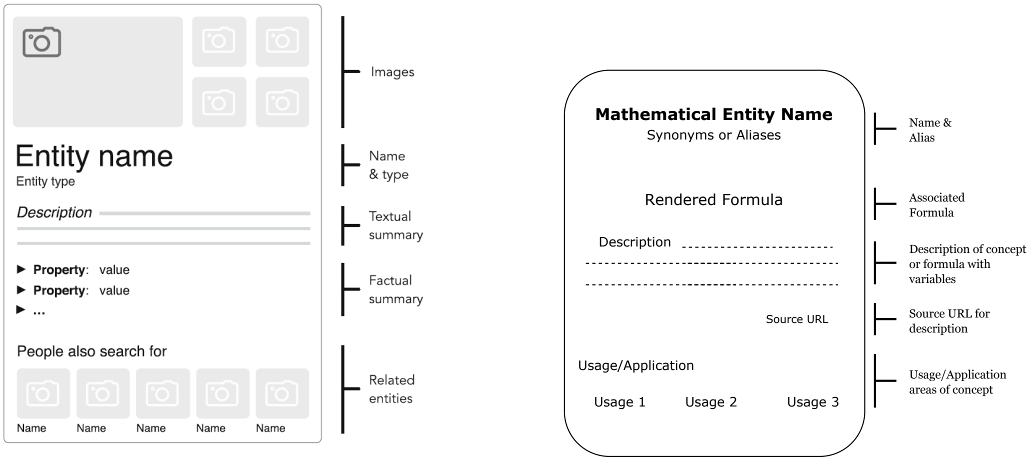
As seen, there is a lack of a sufficiently large annotated data-set to train a machine learning model to identify formulas and their associated definitions in unstructured data. This could be attributed to the difficulty of simultaneously considering the semantics of formulas along with the semantics of the surrounding text while annotating the data. We make use of the earlier approaches in annotating candidates but reduce our candidate pool by considering only structured and semi-structured data known to be concise, thus reducing the uncertainty of whether the text is a description or not. We make use of Wikidata (structured), Wiktionary and ProofWiki (semi-structured) to first identify formulas and then select descriptions and definitions surrounding the formula. Since these data sources, describe a single concept per page/entry disambiguation of the title/name of the mathematical concept is relatively simple.

Chapter 3

Math Entity Card Design

The primary focus of math entity cards are to enable users to navigate seamlessly between formulas and their concepts. By this we mean, allowing users to enter a name of a concept and find its defining formula, or enter a formula and find concepts with which this formula is associated. Entity cards across different commercial text retrieval engines appear to follow a standard design guideline as shown in Figure 3.1a. Users of these search engines have overtime learned to consume a variety of information in the same info-box layout. We wish to use, this familiarity with respect to consuming information in the same layout to our advantage.

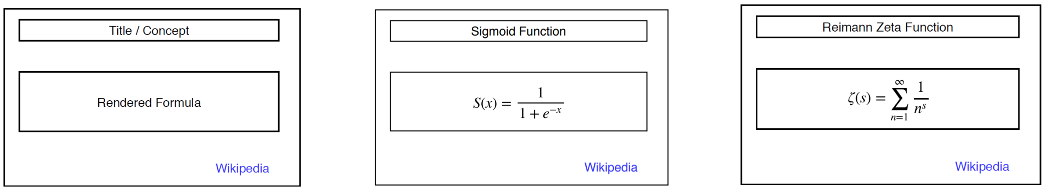
In this chapter, we propose our design decisions for math entity cards, but for the human experiment we make use of the card with only the title, formula and a single description. We propose the addition of a formula field, along with multiple descriptions to support understanding of mathematical concept across different levels of understanding. We also propose the introduction of a usage section that could include examples of the usage or application of the mathematical concept or formula. We introduce math entity cards for symbols, with each card representing a unique concept/functionality for the symbol. We demonstrate the use of math entity cards as a form of auto-complete where in users could enter either the formula or the title of a concept and receive a card directly at query time.



(a) Common Entity Card Layout (b) Math entity card layout

Figure 3.1: Similarities & differences in layout between common entity card as described by Balog K. [2] and proposed math entity card.

3.1 Formula Description Card Designs

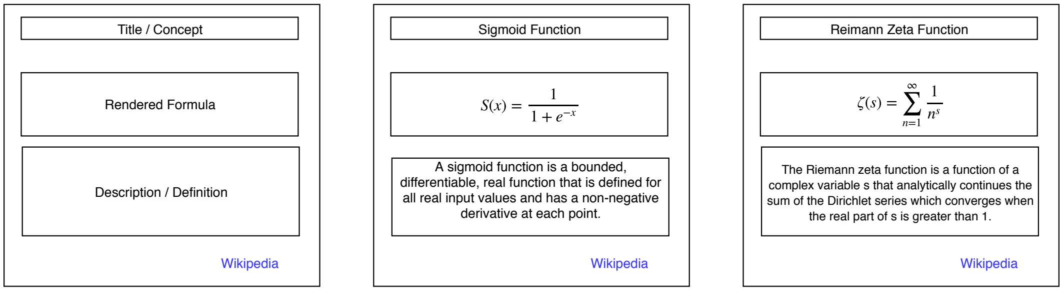


(a) General Template (b) Sigmoid Function (c) Riemann Zeta

Figure 3.2: Examples of math entity cards with title and formula only. Wikipedia indicates the source URL.

Figure 3.2 shows examples of a basic math entity card. We decide to preserve the title and propose to replace the image section with a field for the formulas associated with a math entity. The reason for this is we believe not all mathematical entities can be represented by an image, but they would most likely have a defining formula. We place the formula field just below the title to enable a visual connection between the two. This choice is made keeping in mind that in a math-aware Engine, a user's search would revolve more around formulas and it would be beneficial to have the title and formula as a pair more easily readable. To this basic card design we add a description section (summary) that includes the description of the mathematical concept. Wikipedia acts as the source URL and could point to any source from where the formula/description for the particular mathematical concept is extracted. For our research we consider three data sources, Wiktionary, Wikipedia and Proof Wiki in increasing order of formal descriptions. We believe that due to the complexity of mathematics in general, it is not always feasible to grasp the meaning from one definition and thus having multiple definitions might help. This could also help the more experienced users understand the concept without dilution of information. Also there are some formulas/symbols for example 'α' that are associated with multiple different concepts, in statistics to denote significance level, in machine learning to denote learning rate or angular acceleration in physics. Hence the more varied sources considered, the better our chances at covering multiple concepts.

Three different card designs are presented in increasing amount of information, this is done to analyze how beneficial is mathematical information when summarized and presented in the form of an info-box. The minimal card design in Figure 3.2 presents only the concept name along with the formula that relates to this concept. We noticed during extraction, some formulas have a passing reference of a concept without a description, in such situations it could be at least helpful to provide the user with a name of the concept. This minimal design might suffice in some cases. Users could further decide whether they require additional information and search accordingly with the help of the name of the concept(Title). Sometimes however a description of the formula is needed and supplements the understanding further, as shown in Figure 3.3.

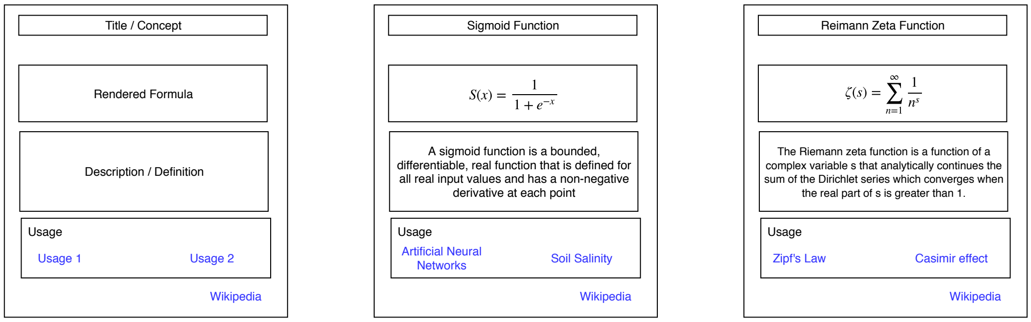

Title / Concept
Rendered Formula
Description / Definition
Wikipedia
Sigmoid Function
$S(x) = \frac{1}{1+e^{-x}}$
A sigmoid function is a bounded, differentiable, real function that is defined for all real input values and has a non-negative derivative at each point.
Wikipedia
Reimann Zeta Function
$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$
The Riemann zeta function is a function of a complex variable s that analytically continues the sum of the Dirichlet series which converges when the real part of s is greater than 1.
Wikipedia

(a) General template (b) Sigmoid Function (c) Riemann Zeta

Figure 3.3: Examples of math entity cards with title-formula and descriptions/definitions

3.2 Math Entity Card: Additional Usage Section

We propose introducing a "Usage" section to indicate other areas where a mathematical concept/formula is used, e.g. Figure 3.4b where a Sigmoid Function is used in Artificial Neural Networks, or the applications of the math



A sigmoid function is a bounded, differentiable, real function that is defined for all real input values and has a non-negative derivative at each point

The Riemann zeta function is a function of a complex variable s that analytically continues the sum of the Dirichlet series which converges when the real part of s is greater than 1.

(a) General template (b) Sigmoid Function (c) Riemann Zeta

Figure 3.4: Examples of math entity cards with title-formula-description and a Usage section

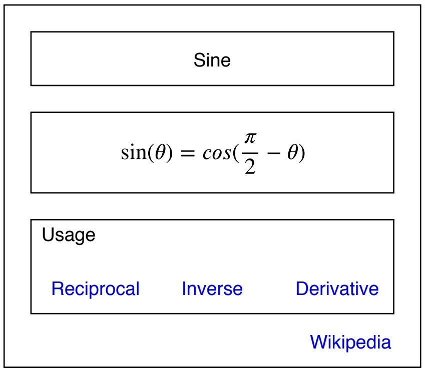


Figure 3.5: $\sin\theta$ card with related functions/operations as usage

concept to other areas as seen in figure 3.4c. The usage area could alternatively be used to include a variety of operations that could be applied to the main function for example Fig. 3.5 where in the user's input query of '$\sin\theta$' results in an Entity Card of $\sin\theta$, instead of the usage however there are three links that describe mathematical operations or transformations that could be applied to the input query. Ideally they should have the following functionality:

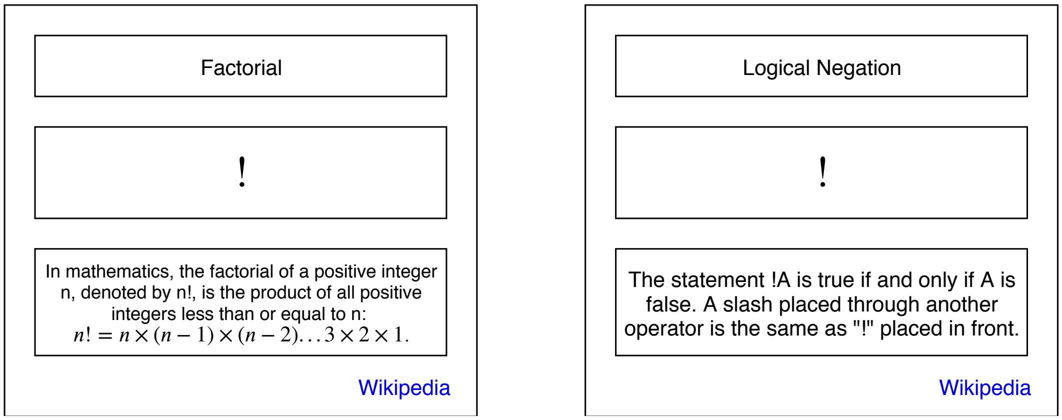
- Reciprocal should lead a user to $\csc\theta$

- Inverse should lead a user to $\theta = \arcsin(\frac{opposite}{hypotenuse})$

- Derivative lead a user to $\cos\theta$

While beneficial, this would require additional research. We believe one way this could be achieved is by fetching sub sections from a Wikipedia page, for example the Wikipedia page for Sine [1] has 'Reciprocal','Inverse' and 'Calculus' as sub sections within Identities, with the help of some text processing it might be possible to fetch meaningful related content. An alternative approach would be to use a system similar to a computer algebra system that can fetch other mathematical concepts that have a relationship with Sine.

3.3 Math Entity Cards for Symbols



In mathematics, the factorial of a positive integer n, denoted by n!, is the product of all positive integers less than or equal to n: $n! = n \times (n-1) \times (n-2) \ldots 3 \times 2 \times 1$.

The statement !A is true if and only if A is false. A slash placed through another operator is the same as "!" placed in front.

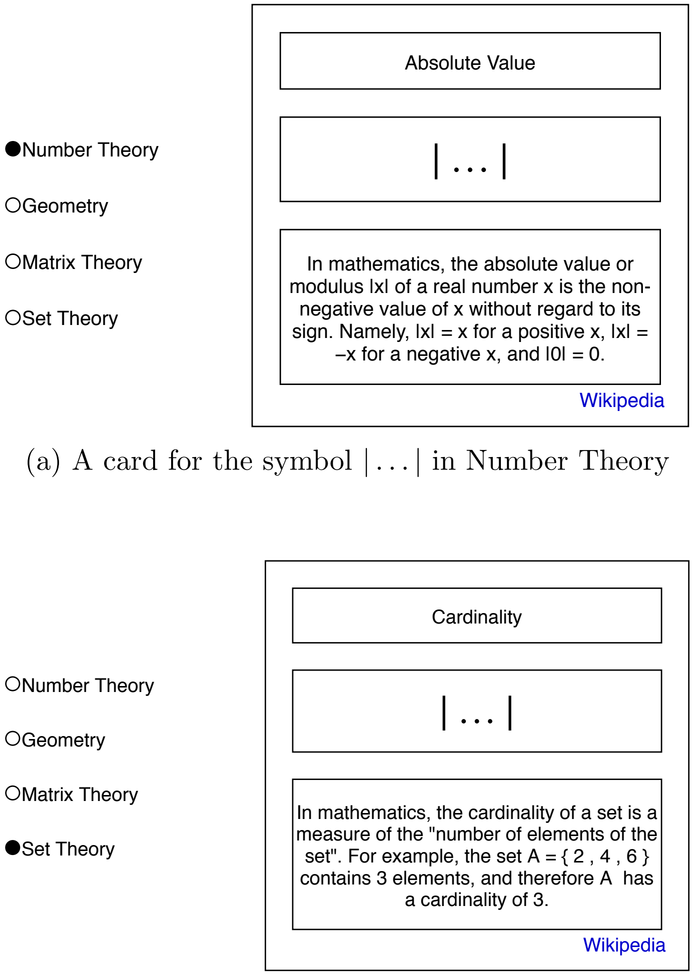
(a) Factorial

(b) Logical Negation

Figure 3.6: Math entity cards for mathematical symbols

[1] https://en.wikipedia.org/wiki/Sine

Formulas are created by a combination of symbols and variables in a manner to convey some meaning or represent a relationship between them. Symbols can hence be considered as independent building blocks of a formula. The template for a math entity card is designed to accommodate math symbols information as well. Users can thus obtain a description of what a symbol represents and know the context in which it is used. This would help reduce the guess work in searching for a symbol. Some symbols are polysemic in nature, i.e., they have multiple meanings, depending on the context in which they are used. For example, **'!'** can be assumed to be either the 'factorial' or 'logical negation' depending on whether a user is concerned with the field of combinatorics or propositional logic. Although the symbol is identical, the concept is different. Hence, we decide to create a new card for every concept attached to a symbol, if they are from different mathematical fields. This opens up the possibility for a search engine to help users narrow down search results by applying multiple filters based on faceted classification of the items (faceted search) as seen in Figure 3.7. These categories (facets) are available for all symbols we extract from the Wikipedia data source.



● Number Theory

○ Geometry

○ Matrix Theory

○ Set Theory

Absolute Value

$|\ldots|$

In mathematics, the absolute value or modulus |x| of a real number x is the non-negative value of x without regard to its sign. Namely, |x| = x for a positive x, |x| = −x for a negative x, and |0| = 0.

Wikipedia

(a) A card for the symbol $|\ldots|$ in Number Theory

○ Number Theory

○ Geometry

○ Matrix Theory

● Set Theory

Cardinality

$|\ldots|$

In mathematics, the cardinality of a set is a measure of the "number of elements of the set". For example, the set A = { 2 , 4 , 6 } contains 3 elements, and therefore A has a cardinality of 3.

Wikipedia

(b) Another card for the same symbol $|\ldots|$ in Set Theory

Figure 3.7: Faceted Search for Symbol Cards

3.4 Alternate Descriptions for a Concept or Formula

As discussed in the introduction of this chapter, we make use of more than one source for information extraction, this was primarily to address the varying information needs for both beginner and intermediate users for the

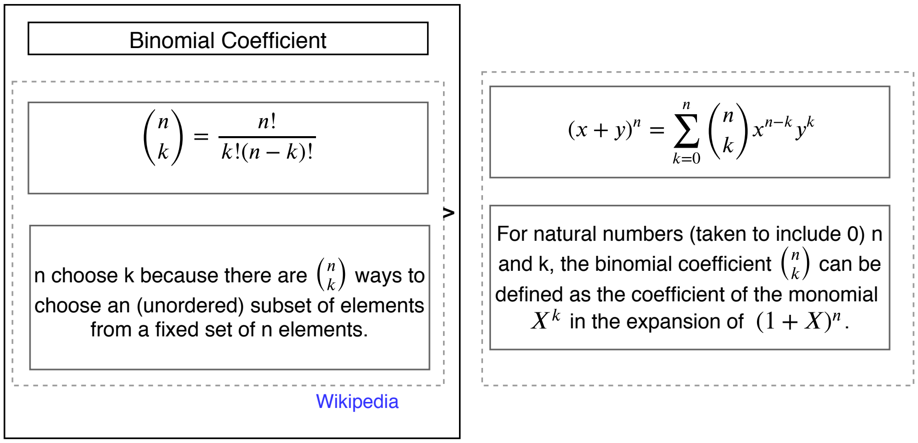
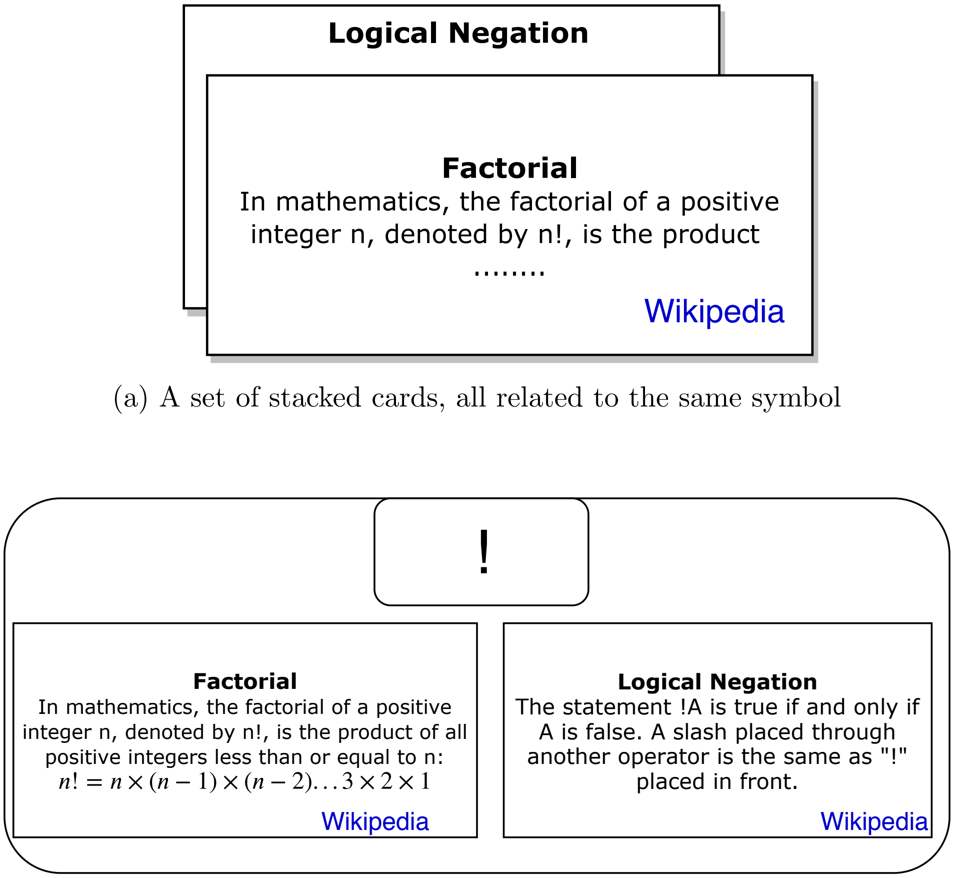

Binomial Coefficient
$\binom{n}{k} = \frac{n!}{k!(n-k)!}$
n choose k because there are $\binom{n}{k}$ ways to choose an (unordered) subset of elements from a fixed set of n elements.
Wikipedia
>
$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$
For natural numbers (taken to include 0) n and k, the binomial coefficient $\binom{n}{k}$ can be defined as the coefficient of the monomial X^k in the expansion of $(1+X)^n$.

Figure 3.8: Binomial Coefficient with multiple formulas and and multiple Descriptions

same mathematical concept. A single mathematical concept can have more than one formula by which it can be identified. For example, in Fig. 3.8 we see the mathematical entity 'Binomial Coefficient' to have more than one possible description. The first describes the way of computing $\binom{n}{k}$, whereas the second describes the occurrence of $\binom{n}{k}$ as part of a broader concept. Either of the descriptions could be beneficial to a user depending on the information need. However each description is closely associated with its individual formula, we refer to this as a Formula-Description pair. Alternatively we could also have multiple formulas but just have a single description associated to the concept in general. To handle multiple formulas with a single description and multiple formula-description pairs with a common presentation, we propose two alternatives:

- Carousel: A carousel feature would enable users to swipe across definitions and formulas treated as pairs. In instances when a concept being searched for has more than one formula-description pair associated with it, the search engine must first display the formula that closely matches the query and then display the others, associated for the concept. This would allow users to continue browsing other formulas connected to the same concept.



In mathematics, the factorial of a positive integer n, denoted by n!, is the product ....

(a) A set of stacked cards, all related to the same symbol

In mathematics, the factorial of a positive integer n, denoted by n!, is the product of all positive integers less than or equal to n: $n! = n \times (n-1) \times (n-2)\ldots 3 \times 2 \times 1$

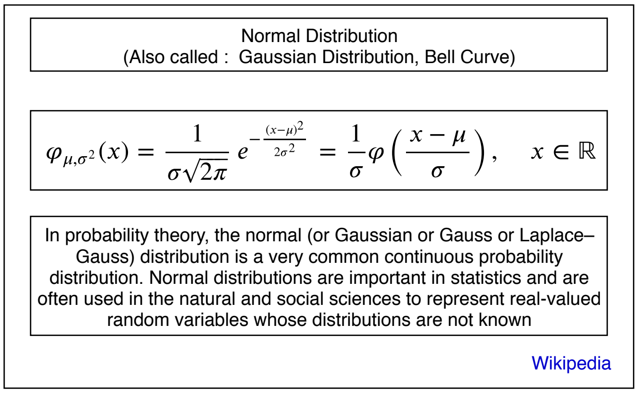
The statement !A is true if and only if A is false. A slash placed through another operator is the same as "!" placed in front.

(b) A Pop up modal interface to view multiple descriptions for symbols.

Figure 3.9: Stacked Cards with Pop Up

- Pop up modal: Most often, we might received multiple formulas or multiple descriptions associated with the concept but not as Formula-Description pairs. In such situations we would need two independent carousels one for the formulas and one for the descriptions. A pop up modal that appears on an action (double-click) instead could help provide multiple snippets of both formulas and/or descriptions that exists for the same concept. This approach provides a more focused view of the mathematical concept being searched for. This approach can also be used for polysemic symbols where the symbol representation stays constant but the Title and Description are displayed as individual components within the pop up as shown in Figure 3.9

3.5 Concept Titles & Aliases



Normal Distribution
(Also called : Gaussian Distribution, Bell Curve)

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}$$

In probability theory, the normal (or Gaussian or Gauss or Laplace–Gauss) distribution is a very common continuous probability distribution. Normal distributions are important in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known

Wikipedia

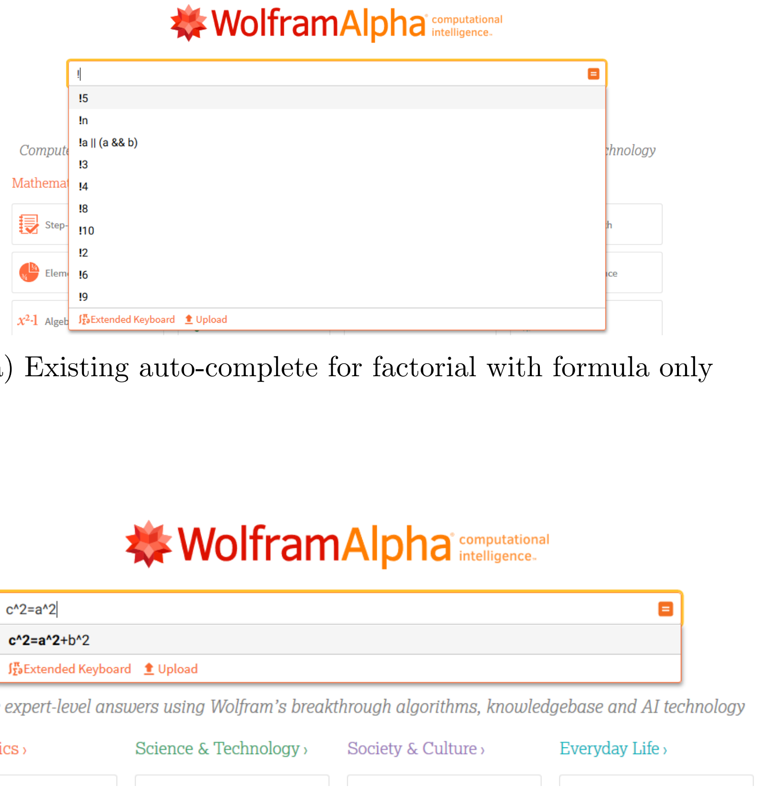
Figure 3.10: Normal Distribution Card with Aliases

There are instances where concepts have multiple different names but have the same formula to represent them. These alternate names acts as synonyms for the concept. In such situations different users might enter different concept names not knowing they all refer to the same concept. These extra synonyms could be used as aliases (alternate names) for the concept. An alias would allow different users to reach the same concept as well as learn the alternative names by which that particular concept can be called. Aliases (i.e. also called) are present in Health Cards generated by Google and Bing, as seen in the study by Jimmy et al. [14]; we propose to make use of the same design (Figure 3.10) to have aliases for math entity cards as well.

3.6 Math Entity Cards in a Math Aware Search Interface

Text based search engines have entity cards as secondary sources of information that are displayed along other results on the Search Engine Result Page (SERP). However for math aware search engines, there is a possibility of providing cards directly at query time as a form of auto-complete, this not only helps save search time but also enhance a user's ability to interact with the search system.

As seen in Figure 3.11a Wolfram Alpha has multiple suggestions for '!' but all of them are different examples of the same concept of factorial rather than showing different concepts where in the '!' exists e.g., Factorial and Logical Negation. This indicates a popularity based ranking, based on past searches or query logs, which is beneficial but does not provide any conceptual information to a user. In Figure 3.11b there is only one suggestion for the formula $c^2 = a^2$, which again confirms a popularity based suggestion. However there most likely exists multiple formulas that overlap with $c^2 = a^2$. This form of popularity based auto-complete of only formulas has two limitations; one it only provides formulas and no other information regarding the formula and second it could be easily affected, if multiple users query the same formula with minor changes to either the variables or the order of operations, this makes searching for new concepts that have an overlap with the formula, difficult for a user.



!

c^2=a^2

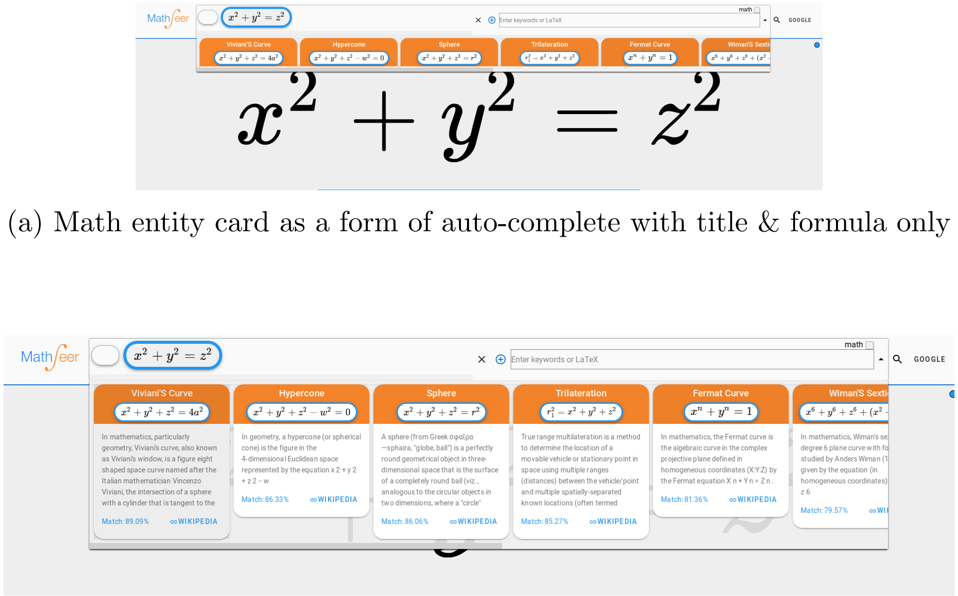
(a) Existing auto-complete for factorial with formula only

(b) Existing auto-complete for $c^2 = a^2$ with only one match

Figure 3.11: LaTeX formula auto-complete as present in Wolfram Alpha

Math entity cards help provide auto-complete results based on mathematical entities to which formulas are associated, this is not affected by a popularity based search. As shown in Figure 3.12a by providing the title and the formula when users enter a formula, users are given conceptual feedback of a formula. This helps a user relate to concepts immediately, we assume this would be more beneficial as compared to just providing other formulas that appear similar since text is more prevalent than formulas. By indexing both the formula and the title, a user could enter either a partial formula or partial title to browse other math concepts having formulas/titles that overlap with the input query. This could serve a purpose of comparison based decision making, similar to the multi-card interface proposed by Jimmy et al. [15]. By clicking on the card users are able to receive descriptions for concepts, that provides information directly, this could either satisfy a factual information need, or help alter a navigational information need.

Given the development of multi-modal canvas based systems such as *min* [31] and their demonstrated usefulness in drawing and editing formulas [33, 35], it would be beneficial for a user to drag the formula from an entity card on to the canvas, modify it and search for other diverse results from there.



(a) Math entity card as a form of auto-complete with title & formula only

In mathematics, particularly geometry, Viviani's curve, also known as Viviani's window, is a figure eight shaped space curve named after the Italian mathematician Vincenzo Viviani, the intersection of a sphere with a cylinder that is tangent to the

Match: 89.09% WIKIPEDIA

In geometry, a hypercone (or spherical cone) is the figure in the 4-dimensional Euclidean space represented by the equation x 2 + y 2 + z 2 − w

Match: 86.33% WIKIPEDIA

A sphere (from Greek σφαῖρα —sphaira, "globe, ball") is a perfectly round geometrical object in three-dimensional space that is the surface of a completely round ball (viz., analogous to the circular objects in two dimensions, where a "circle"

Match: 86.06% WIKIPEDIA

True range multilateration is a method to determine the location of a movable vehicle or stationary point in space using multiple ranges (distances) between the vehicle/point and multiple spatially-separated known locations (often termed

Match: 85.27% WIKIPEDIA

In mathematics, the Fermat curve is the algebraic curve in the complex projective plane defined in homogeneous coordinates (X:Y:Z) by the Fermat equation X n + Y n = Z n .

Match: 81.36% WIKIPEDIA

In mathematics, Wiman's se degree 6 plane curve with fo studied by Anders Wiman (1 given by the equation (in homogeneous coordinates) z 6

Match: 79.57% WI

(b) Math entity card as a form of auto-complete with title-formula & Description

Figure 3.12: Math entity Cards as auto-complete

3.7 Summary

We have seen alternate designs for math entity cards and the design decisions that make them better suited for math information retrieval. We have also briefly discussed applications of math entity cards in this chapter. We will now discuss the extraction methods to populate each section of the math entity cards.

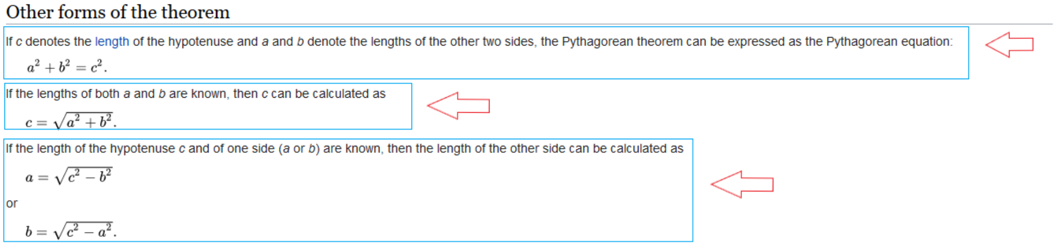
Chapter 4

Math Entity Card Creation

As seen in the related work, extracting descriptions and titles for a formula from unstructured data requires manual annotation and could introduce a class imbalance problem. Given the presence of massive online open source knowledge bases such as DBPedia, Wikidata, Wiktionary and ProofWiki we decide to make use of basic data processing and rule-based information extraction techniques to create math entity cards. We first describe the creation of math entity cards for the purpose of using them within MathSeer, a math-aware search interface. These cards have additional features (alternate descriptions and keyword based search) that are beneficial to have, but do not yet have any formal experiment to confirm their benefits. Hence for our human experiment we created cards without the additional features.

4.1 Math Entity Card Creation

In regular Wikipedia or scientific articles, a single page can contain more than a single formula, this requires us to solve both an Entity Identification task (which formula amongst the others on the page represents the concept) and Information Extraction (fetching a valid description of that for-



Other forms of the theorem

If *c* denotes the length of the hypotenuse and *a* and *b* denote the lengths of the other two sides, the Pythagorean theorem can be expressed as the Pythagorean equation:

$$a^2 + b^2 = c^2.$$

If the lengths of both *a* and *b* are known, then *c* can be calculated as

$$c = \sqrt{a^2 + b^2}.$$

If the length of the hypotenuse *c* and of one side (*a* or *b*) are known, then the length of the other side can be calculated as

$$a = \sqrt{c^2 - b^2}$$

or

$$b = \sqrt{c^2 - a^2}.$$

Figure 4.1: A section of Pythagorean Theorem from Wikipedia, highlighted are multiple valid definitions.

mula and concept). Figure 4.1 is a section of the Wikipedia page for the Pythagorean Theorem[1] and demonstrates an example of the challenges faced for fetching title, formula and description of a mathematical concept from unstructured data source such as Wikipedia.

We decide to simplify the process of card creation to focus first on understanding the usefulness of math entity cards. We hence resort to fetching data from sources that reduce the ambiguity between multiple formulas on a page and the number of valid descriptions of the concept. We begin by extracting information from Wikidata (a structured knowledge base) and then supplement it with information from Wiktionary and ProofWiki (semi-structured knowledge bases). Wikidata is a structured knowledge base that has an entity relationship (defining formula) as a specific field for mathematical formulas, and while its is possible to find individual pages with the help of a URI[2] and QID, it is not currently possible to query for the formulas directly.

[1] https://en.wikipedia.org/wiki/Pythagorean_theorem

[2] http://www.wikidata.org/entity/

We hence first query Wikidata via its SPARQL end point[3] for all the entries that have a formula. This helps fetch formula, titles, descriptions and aliases if any directly. Wikitionary and ProofWiki are considered as semi-structured since each of the sources, have definitions for the concept demarcated under specific section headers (e.g., ProofWiki == Definition ==). However, each of these definitions could have more than one formula and there is a need to find which formula should be associated with the concept. We extract, clean and processes data from each of the available data stores and store them in our relational knowledge base designed specifically for math entity card creation.

4.1.1 Extracting Formulas & Titles From Structured Data Sources

Wikidata is a structured representation of Wikipedia. Its data is available for download in JSON, RDF, and XML formats, and can be access via a search API[4]. Each entry in Wikidata can be uniquely identified by an id, also called QID, or Wikidata QID. Every entry has individual property identifiers such as ISBN-13 (P212) that identifies books, or producer (P162) that identifies person(s) who produced a film, musical work or other art works. Mathematical Entities have a defining formula (P2534) property by which they can be identified, which are represented in presentation MathML [1] format. As of July 2019, 3644 entities were discovered that had at least one mathematical formula. Out of the 3644 entries, 35 entries have a mathematical formula

[3] https://query.wikidata.org/

[4] http://www.wikidata.org/entity/

but only have their corresponding QIDs in the title, which does not convey any useful information and are hence omitted. We found one duplicate entry with difference in letter case, 'First Law of Thermodynamics' (Q25209772) vs 'first law of thermodynamics ' (Q179380). They both however have different formulas, which we save under a single entry (First Law of Thermodynamics), this gives us 3608 unique concepts having at least one mathematical formula. Table 4.1 shows the distribution of formulas per concepts, as we see a considerable majority of concepts have a single formula. A total of 3572 mathematical formulas were identified for the 3609 Concepts. Of these, 6 formulas were identified as mathematical symbols. Some of these formulas have references that point a user to the source from where they are obtained, the others do not. On checking the references we realize that Wikidata internally has bots importing data from Wikimedia projects. This means the associations between formulas and concepts are not always correct, and need some manual cleaning. But on a visual inspection of 10% of the data, they seem accurate enough to be used for math entity card creation. Further data validation is beyond the scope of this thesis, as it would require manually checking every association. Existing entity cards in search engines circumvent this problem by adding a feedback option below the entity card, allowing users to provide feedback and point out the ones that are incorrect

Wikidata also has an 'itemDescription' property that could we use to fill in the description section of a math entity card. However, we found roughly 56% (2038/3608) of the records have no description, and another 16% (577/3608)

Table 4.1: Distribution of Formulas per Concept

Formulas per Concept	1	2	3	4	5	6	11
Number of Entries	3495	98	9	3	1	1	1

have less than five words in the description field. The reason for selecting five words as a threshold, was to avoid descriptions containing single words or incomplete sentences such as “Algorithm”, “Image Processing”,“Theorem”,“Irreducible Fraction”. All descriptions from Wikidata are saved, but ranked lower in order while displaying on the entity card, this is to support multiple descriptions for a mathematical entity. For descriptions, we keep the following order based on the technicality of the language in the description: Wikitonary (least formal), Wikipedia, Wikidata and ProofWiki (most formal). Wikidata also provides us with alternate names or aliases under the ‘itemAltLabel’ property. Using this we extract 827 aliases for the 3609 concept titles.

4.1.2 Adding Mathematical Concept Descriptions Using Wikipedia

We query and fetch the first two sentences on a page to fetch more meaningful & complete descriptions mainly for the 72% of Wikidata concepts (with descriptions that are empty or have less than five words). Every page in Wikipedia is structured such that the opening paragraph (also called ‘Lead Section’) is general in style, and serves as an introduction to the article as a whole [5]. We extract the first two sentences inclusive of any math expressions present in them. Existing entity cards for text search engines decide to either

[5] https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Lead_section

extract these math expressions as text (Bing & DuckDuckGo) or skip them altogether (Google). Given the effect, that rendered mathematical expressions have in relevancy assessment as studied by Reichenbach et al. [29] we render the expressions within the card description sections as well.

4.1.2.1 Extracting Symbol Content

Since Wikidata has only 6 mathematical symbols, we decide to extract more symbol information from Wikipedia. 'List of Mathematical Symbols' [6] is a dedicated page for symbols with the components (Symbol in Tex, symbol, name, and description) needed to create a symbol card. Each symbol also has two additional components, 'category' which could be used as tags to enable faceted search and an 'explanation' column that provides examples for each meaning of the symbol. We are able to fetch 209 Title - Symbol - Description triples, for a total of 187 unique symbols.

4.1.3 Extracting Formula, Title & Defintions From Wiktionary

Wiktionary is a multilingual, web-based project to create a free content dictionary of terms in all natural languages. The coverage of mathematical formulas is not extensive, with most definitions missing the associated formulas. In Wiktionary the language used is less formal/technical and hence easier to read and understand. Also, unlike Wikidata the descriptions are more complete in sentence structure. This is mainly due to the fact that Wiktionary is

[6] https://en.wikipedia.org/wiki/List_of_mathematical_symbols

designed mainly to be used as a dictionary.

Wiktionary's data like other Wikimedia projects [Wikipedia, Wikibooks, Wikiquote, Wikiversity] make their data available as an XML dump[7]. Wikitionary is a multilingual data-source, but for the current version of math entity cards, we fetch and process only the English version of the dumps (prefixed with 'enwiktionary'). As with most of Wikimedia data, the dumps have the data in wiki markup[8] format stored within XML. Not all Wikitionary pages have math content. We first filter those Wikitionary pages that have any mathematical content with the help of regex matching, searching for '&lt;math&gt' within the text content of the body. From this we filter wikitionary internal pages (having titles:wiktionary:tea_room, wiktionary-:information_desk, wiktionary:etymology_scriptorium etc.). Eight english content pages have their titles written in Chinese Characters, we filter these out as well, thus resulting in a total of 1376/6571189 that have some math content. With the help of Pandoc[9] we convert these pages to HTML format for easier processing. 507 pages however lose their mathematical content on conversion via Pandoc without any error while conversion. This results in a total of 861 pages that we use to create cards from Wiktionary.

We use a two pass-approach to extract the content. On the first pass we pick those text paragraphs that have at least one mathematical expression and have a match percentage greater than 70% of the respective page title in a

[7] https://dumps.wikimedia.org/backup-index.html
[8] https://en.wikipedia.org/wiki/Help:Wikitext
[9] https://pandoc.org/

Table 4.2: Match percentage between title and strong tag contents in Wiktionary

Page Title	**Strong Tag Content**	**Match % of Title and Strong Content**
algebraic number	algebraic numbers	97
pauli matrix	pauli matrices	85
group theory	group theories	87
σ-algebra	sigma algebra	73
well-order	well orders	86

Table 4.3: Number of Math Per Description Wiktionary
Note : Titles are not exclusive, some titles have multiple descriptions

Number of Formulas per Description	1	2	3	4	5	6	7	8	9	10
Number of Titles	281	172	104	77	49	41	30	24	16	12

'strong' tag. We decide to use the SequenceMatcher class from the difflib package, that implements the Gestalt Pattern Matching algorithm, the algorithm does not yield minimal edit distances, rather yields matches that "look right to people." This approximate matching is done against the page title rather than exact matching to account for plural and minor differences that have no change in meaning. Table 4.2 shows some examples of the approximate matching. We collect a total of 336 Title-Description pairs via this approach. The second pass is done for those pages that do not have any strong elements in the paragraph containing math. We extract the first text description that follows either of the following header ids ('numeral', 'adjective', 'noun', 'symbol', 'proper-noun'). This results in an additional 300 title-description pairs. We thus extract a total of 636 unique title-description pairs from 861 pages and then proceed to selecting a single math expressions to be associated with each description.

4.1.3.1 Selecting A Single Math Expression From Wikitionary Definition

Every description in Wiktionary has at least one mathematical expression, but some have more. Table 4.3 shows the distribution of math Expressions by the Number of Titles. Math entity cards, have a single title but can have multiple formulas and multiple descriptions. However every formula or description should be associated with only that mathematical entity. This is to avoid random mathematical expressions showing up as the formula related to a mathematical Entity. The descriptions having a single math expressions are extracted as is, with the expression being the main formula. For the others we use verbal cues and pick the math element that follows the strong element, this is similar to the approach used by [28] for extracting Concept-Formula-Description Triples. To avoid a large number of flase positives (math element selected but are not representative of the concept) we only make use of the above two rules, giving us a total of 483 unique concepts with corresponding formulas and descriptions.

4.1.4 Extracting Formal Mathematical Definitions From ProofWiki

ProofWiki is described as "an online compendium of mathematical proofs." Their goal is the collection, collaboration and classification of mathematical proofs. As of date they have 17,954 Proofs & 13,894 Definitions. The language in ProofWiki is relatively more formal compared to Wikidata or Wiktionary. We noticed however the ProofWiki is not exhaustive as a data set and there were some mathematical concepts that did not have definitions which were present in Wiktionary (e.g., Sigmoid Function).

ProofWiki has a separate namespace for definitions that helps categorize the data, however it also makes use of the template based wiki markup[10] syntax, that prevents us from extracting the definitions directly. We hence first crawl through the entire collection of Definitions and create a dictionary based mapping of the main pages and its sub-pages. For example the definition page for Binomial Coefficient (Definition:Binomial Coefficient) pulls content from the following sub-pages :

- Definition:Binomial Coefficient/Integers/Definition 1
- Definition:Binomial Coefficient/Integers/Definition 2
- Definition:Binomial Coefficient/Integers/Definition 3
- Definition:Binomial Coefficient/Real Numbers
- Definition:Binomial Coefficient/Complex Numbers
- Definition:Binomial Coefficient/Multiindices
- Definition:Binomial Coefficient/Notation
- Definition:Binomial Coefficient/Historical Note
- Definition:Binomial Coefficient/Technical Note

[10] https://en.wikipedia.org/wiki/Help:Wikitext

$$\text{Definition:Binomial Coefficient}\left\rangle\begin{array}{l}\text{/Integers/Definition 1}\\ \text{/Integers/Definition 2}\\ \text{/Integers/Definition 3}\\ \text{/Real Numbers}\\ \text{/Complex Numbers}\\ \text{/Multiindices}\\ \text{/Notation}\\ \text{/Historical Note}\\ \text{/Technical Note}\end{array}\right.$$

By this we receive 8919 unique page headers, having a total of 5385 sub pages not including the header page (Definition:Binomial Coefficient). Some main pages have definitions, where as the other fetch content from subpages present within <onlyinclude>and <\onlyinclude> is fetched, we do the same and fetch content between the first header of definition (== Definition==) and any immediate next header, this style of ProofWiki makes the extraction process simple. We then check for the presence of any 'onlyinclude' tags and if present fetch content within the tags. We skip pages that have either 'Notation' or 'Note' in the title e.g Historical Note or Techincal Note, this is done since although they have content within (== Definition ==) header the content is not currently useful in math entity cards. We end up with a total of 9279 definitions.

To ensure, that the fetched definition has a formula for the concept, we filter only those definitions that have math, giving us a total 7428 definitions. With the help of Pandoc[11] we convert these pages to HTML format for easier

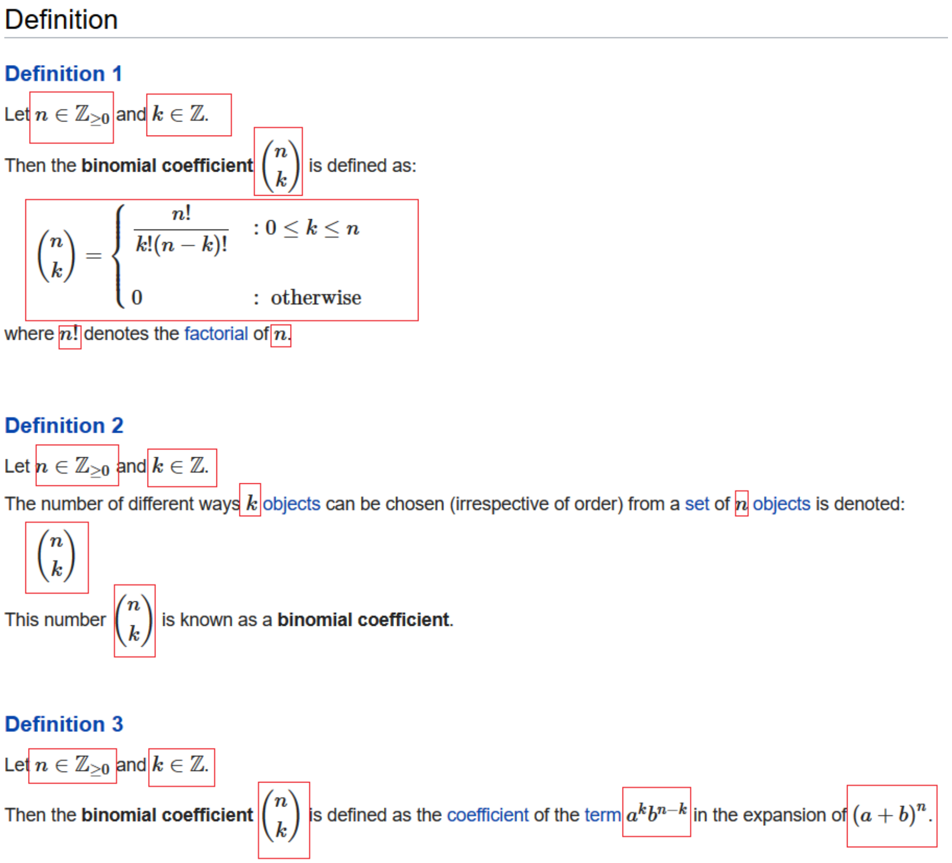
[11] https://pandoc.org/

processing. Ten pages have an error while converting with Pandoc leaving us a total of 7418.

4.1.4.1 Selecting A Single Math Expression From ProofWiki Definition

As seen in Figure 4.2, there are multiple math formulas within each definition on a page. Each math is represented in LaTeX format within '\$\$' signs. To extract the math from each definition, we make use of context and language cues present in the inherent nature of ProofWiki definitions. We break up the definition into sentences and process each sentence to first check for the presence of math. If a sentence has math, we apply the following handcrafted rules to extract formulas.

1. Sentence starts with 'Let': Skip sentence
2. Sentence has strong element and sentence has colon: Get formula after colon
3. Sentence has keywords (defined, denoted) and sentence has colon: Get formula after first strong element.
4. Sentence has strong element and sentence has no colon: Get formula after first strong element
5. Sentence has no strong and sentence has colon: Get formula after colon
6. All sentence has only 'Let': Get formula after colon for sentence that has a colon.



Definition

Definition 1

Let $n \in \mathbb{Z}_{\geq 0}$ and $k \in \mathbb{Z}$.

Then the **binomial coefficient** $\dbinom{n}{k}$ is defined as:

$$\binom{n}{k} = \begin{cases} \dfrac{n!}{k!(n-k)!} & : 0 \le k \le n \\ 0 & : \text{otherwise} \end{cases}$$

where $n!$ denotes the factorial of n.

Definition 2

Let $n \in \mathbb{Z}_{\geq 0}$ and $k \in \mathbb{Z}$.

The number of different ways k objects can be chosen (irrespective of order) from a set of n objects is denoted:

$$\binom{n}{k}$$

This number $\dbinom{n}{k}$ is known as a **binomial coefficient**.

Definition 3

Let $n \in \mathbb{Z}_{\geq 0}$ and $k \in \mathbb{Z}$.

Then the **binomial coefficient** $\dbinom{n}{k}$ is defined as the coefficient of the term $a^k b^{n-k}$ in the expansion of $(a+b)^n$.

Figure 4.2: ProofWiki page for Binomial Coefficient with math highlighted in Red Boxes

The rules, when applied in order,help extract the formula that we consider to be the best representative of the concept. For example in Figure 4.2 these rules extract the following formulas for definitions:

- Definition 1, formula extracted $\dbinom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & : 0 \le k \le n \\ 0 & : \text{ otherwise} \end{cases}$
- Definition 2, formula extracted $\dbinom{n}{k}$

Table 4.4: Multiple Descriptions for the Binomial Coefficent from different data sources.

Source	**Description/Definition**
Wikidata	family of positive integers that occur as coefficients in the binomial theorem
Wiktionary	a coefficient of any of the terms in the expansion of the binomial $(x+y)^n$, defined by $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, read as "n choose k"
ProofWiki	Let $n \in \mathbb{Z}_{\geq 0}$ and $k \in \mathbb{Z}$. The number of different ways k objects can be chosen (irrespective of order) from a set of n objects is denoted: $\binom{n}{k}$

- Definition 3, formula extracted $\binom{n}{k}$

With the help of this method we are able to extract formulas for a total of 6774 definitions. The remaining 644 for pages either have math that is not represented by LaTeX or cannot be generalized by rules and would have to be handled on a case by case basis, which would not be consistent and are hence omitted.

4.2 Synthesizing the Data

Since we have data from different data sources, each having a different representation of math formulas (MathML: Wikidata and LaTeX: ProofWiki and Wiktionary) we decide to convert and store all data in LaTeX format. We choose LaTeX since its is more easily inter convertible with the help of other tools such as MathJax[12] or LaTeXML[13]. We make use of SQLite[14] to store the data in relational format. We decided to use a relational format due to the implicit relationships between mathematical Entities, their formulas, descriptions and aliases. Table 4.4 shows an example of the multiple descriptions for the Binomial Coefficient that are present in our data set. We check for overlap between concepts and formulas only, i.e if we find an alternate formula for an existing concept, we create a new entry for the formula and add a reference to the existing concept. If however we find an exact match between an existing formula as measured by TangentCFT[22] and the name of the concept is not a match, we add the new title as an alias for the existing concept.

4.2.1 Math Entity Card Prototype & API for Auto-complete

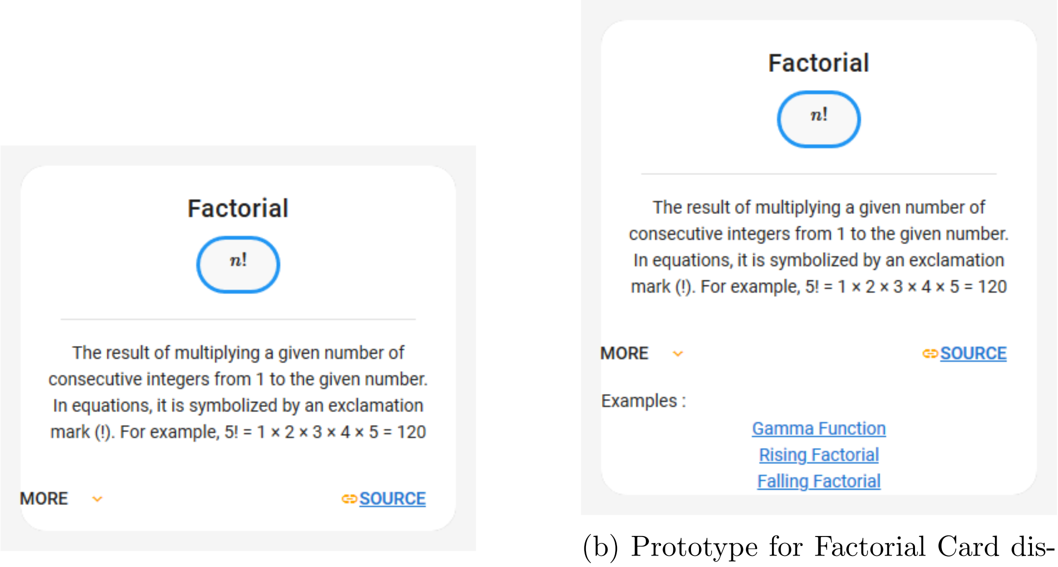
This section describes the card prototype developed in Vue with the help of the Vuetify framework and the REST API for math entity cards.

A prototype for the various functionalities of the math entity card was created using the Vuetify frontend framework. It makes use of static response data that would ideally be returned from a REST API, they include, the title, formula as MathML, (LaTeX could also be used and then rendered in the front end using MathJax), description and corresponding source name and source URL. Figure 4.3 demonstrates the card components with title,

[12] https://www.mathjax.org/

[13] https://dlmf.nist.gov/LaTeXML/

[14] https://www.sqlite.org/



Factorial

$n!$

The result of multiplying a given number of consecutive integers from 1 to the given number. In equations, it is symbolized by an exclamation mark (!). For example, 5! = 1 × 2 × 3 × 4 × 5 = 120

(a) Prototype for Factorial Card with title, formula and description.

Factorial

$n!$

The result of multiplying a given number of consecutive integers from 1 to the given number. In equations, it is symbolized by an exclamation mark (!). For example, 5! = 1 × 2 × 3 × 4 × 5 = 120

Examples :

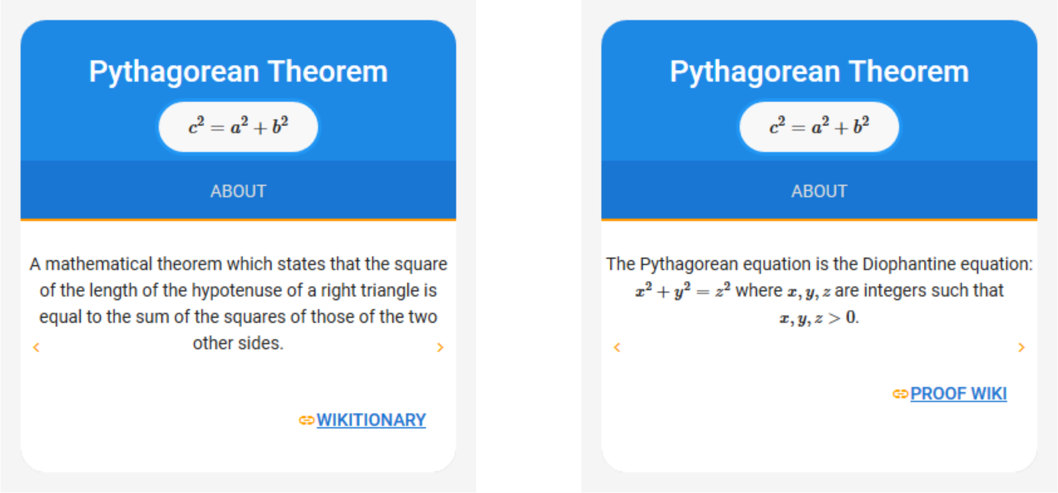
Gamma Function
Rising Factorial
Falling Factorial

(b) Prototype for Factorial Card displaying related concepts on click of more.

Figure 4.3: Prototype of Math Entity Cards

description and formula along with a 'more' section that displays examples such as Gamma Factorial, Rising Factorial, Falling Factorial as hyperlinks. Figure 4.4a and 4.4b demonstrate the example of the rotating carousel feature for multiple descriptions. As seen the descriptions are ordered in increasing order of formality of language, this order is taken directly from the datasources, i.e., Wiktionary, Wikipedia, Wikidata, ProofWiki with Wiktionary being the least formal and ProofWiki being the most formal.

As seen in Figure 4.5, every formula has a Formula ID, which is a unique ID, used as a foreign key to map to its corresponding concept. With the help of TangentCFT [22] formula embedding approach, a formula embedding vector for every formula is created and converted to a BLOB and stored in



Pythagorean Theorem

$c^2 = a^2 + b^2$

ABOUT

A mathematical theorem which states that the square of the length of the hypotenuse of a right triangle is equal to the sum of the squares of those of the two other sides.

‹ ›

WIKITIONARY

Pythagorean Theorem

$c^2 = a^2 + b^2$

ABOUT

The Pythagorean equation is the Diophantine equation: $x^2 + y^2 = z^2$ where x, y, z are integers such that $x, y, z > 0$.

‹ ›

PROOF WIKI

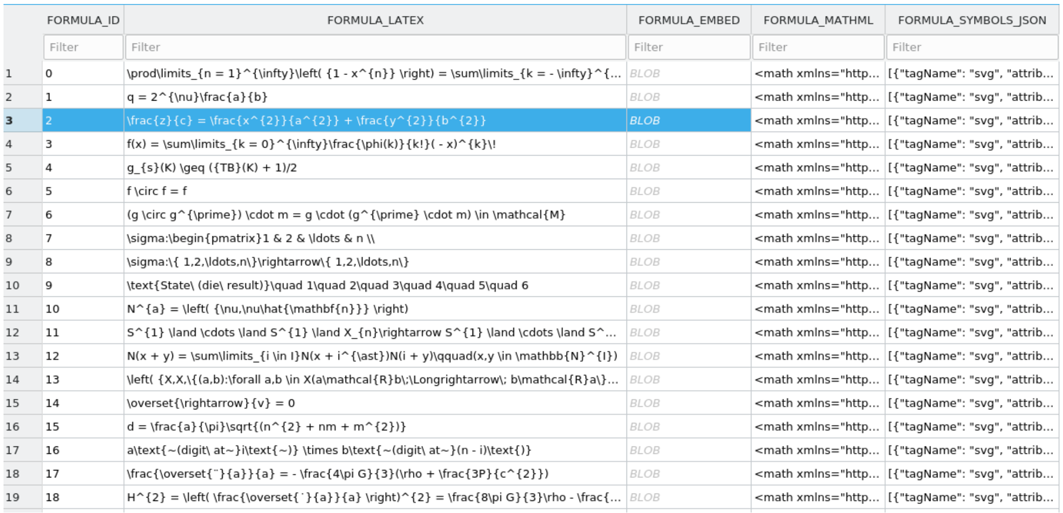
(a) Prototype for Pythagorean Theorem with title, formula and definition taken from Wiktionary.

(b) Prototype for Pythagorean Theorem with title, formula and definition taken from ProofWiki.

Figure 4.4: Prototype of math entity cards with Carousel for rotating descriptions.

the database. For an input query in LaTeX, TangentCFT is again used to convert the input query to a formula vector and rank all the existing formula embedding vectors as per cosine similarity. The top 10 highest matches are selected, since a smaller number would mean more search requests, where as a larger would require an additional search amongst the returned results. Text-search engines also return the top-10 links for a search query. Since we now have the highest matched formula ID, we make use of the foreign key mapping to concepts, and descriptions to then fetch all the data that can be used to populate a math entity card. Along with this any tag or alias data is also returned as part of the JSON response.

The prototype has been modified by Gavin Nishizawa to suit the functionality



	FORMULA_ID	FORMULA_LATEX	FORMULA_EMBED	FORMULA_MATHML	FORMULA_SYMBOLS_JSON
	Filter	Filter	Filter	Filter	Filter
1	0	\prod\limits_{n = 1}^{\infty}\left({1 - x^{n}} \right) = \sum\limits_{k = - \infty}^{...	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
2	1	q = 2^{\nu}\frac{a}{b}	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
3	2	\frac{z}{c} = \frac{x^{2}}{a^{2}} + \frac{y^{2}}{b^{2}}	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
4	3	f(x) = \sum\limits_{k = 0}^{\infty}\frac{\phi(k)}{k!}(- x)^{k}\!	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
5	4	g_{s}(K) \geq ({TB}(K) + 1)/2	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
6	5	f \circ f = f	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
7	6	(g \circ g^{\prime}) \cdot m = g \cdot (g^{\prime} \cdot m) \in \mathcal{M}	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
8	7	\sigma:\begin{pmatrix}1 & 2 & \ldots & n \\	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
9	8	\sigma:\{ 1,2,\ldots,n\}\rightarrow\{ 1,2,\ldots,n\}	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
10	9	\text{State\ (die\ result)}\quad 1\quad 2\quad 3\quad 4\quad 5\quad 6	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
11	10	N^{a} = \left({\nu,\nu\hat{\mathbf{n}}} \right)	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
12	11	S^{1} \land \cdots \land S^{1} \land X_{n}\rightarrow S^{1} \land \cdots \land S^...	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
13	12	N(x + y) = \sum\limits_{i \in I}N(x + i^{\ast})N(i + y)\qquad(x,y \in \mathbb{N}^{I})	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
14	13	\left({X,X,\{(a,b):\forall a,b \in X(a\mathcal{R}b\;\Longrightarrow\; b\mathcal{R}a\}...	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
15	14	\vec{v} = 0	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
16	15	d = \frac{a}{\pi}\sqrt{(n^{2} + nm + m^{2})}	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
17	16	a\text{~(digit\ at~}i\text{~)} \times b\text{~(digit\ at~}(n - i)\text{)}	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
18	17	\frac{\overset{¨}{a}}{a} = - \frac{4\pi G}{3}(\rho + \frac{3P}{c^{2}})	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...
19	18	H^{2} = \left(\frac{\overset{˙}{a}}{a} \right)^{2} = \frac{8\pi G}{3}\rho - \frac{...	BLOB	<math xmlns="http...	[{"tagName": "svg", "attrib...

Figure 4.5: A subset of formulas, along with other fields as stored in the database

of MathSeer, the modifications include changing of the formula to actually behave as a chip and removal of carousel feature. The concept of a chip exists only within the MathSeer interface and hence has not been included in the prototype. The removal of the carousel feature within the card, since the cards are to be used as an auto-complete feature. To enable faster response times, every formula in the database has been pre-rendered into an SVG by a script written by Gavin Nishizawa. The script saves a JSON response of all formula IDs and corresponding SVGs which is then inserted into the database by the author, this is also returned as part of the API response to enable quick response times for MathSeer, the rest of the API is queried as is to serve the auto-complete feature as part of MathSeer. The code for the prototype[15] and

[15] https://gitlab.com/Dmello/math_entity_card_prototype

card database creation and API[16] is available for download

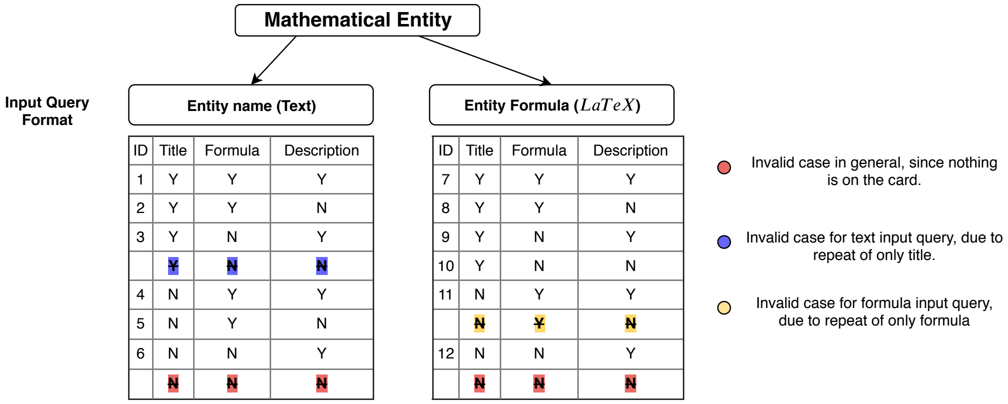
Overall we have been able to extract and synthesize a total of 8870 unique concepts, 1009 aliases, 59 tags, 9681 unique mathematical formulas and 10737 descriptions. We next use a subset of this collection to create math entity cards having only a single description and perform a human experiment to evaluate the usefulness of these cards.

[16] https://gitlab.com/Dmello/math_entity_card

Chapter 5

Human Experiment

We conduct a human experiment to observe any differences, in usefulness of individual card components (title, formula, description) under two scenarios, LaTeX queries and text queries. This helps us observe user preferences of math Entity Cards and understand user's information requirements to help focus future efforts on improvising card contents.



ID	Title	Formula	Description
1	Y	Y	Y
2	Y	Y	N
3	Y	N	Y
	~~Y~~	~~N~~	~~N~~
4	N	Y	Y
5	N	Y	N
6	N	N	Y
	~~N~~	~~N~~	~~N~~

ID	Title	Formula	Description
7	Y	Y	Y
8	Y	Y	N
9	Y	N	Y
10	Y	N	N
11	N	Y	Y
	~~N~~	~~Y~~	~~N~~
12	N	N	Y
	~~N~~	~~N~~	~~N~~

🔴 Invalid case in general, since nothing is on the card.

🔵 Invalid case for text input query, due to repeat of only title.

🟡 Invalid case for formula input query, due to repeat of only formula

Figure 5.1: Different combinations of card components (Title, Formula, Description) forming different card types.

Figure 5.1 shows the six different card types across two query conditions that are considered for a single mathematical entity. We hypothesize that the presence of a math entity card for a LaTeX query would be perceived as more useful as compared to a math entity card for a text query. We assume this mainly since the navigation from text to formula is more common than formula to text. In addition, we wish to observe, whether having prior knowledge of a topic causes any difference in how useful a card is perceived. We also wish to figure out if a single component or a pair of components has the most usefulness in terms of addressing a factual informational need, such as searching for a name/alias, definition, derivation, explanation or application etc [36].

5.1 Experiment Design

A within-subject design is used such that every participant sees a single card type for a mathematical entity. A single participant will see a total 48 unique mathematical entities. As advised by Hearst [10], we decide not to repeat mathematical entities to avoid any learning effect between card contents. Every participant has four practice trials to familiarize themselves with the interface, the responses for the practice trials are recorded but are not used in result analysis. The forty-eight mathematical entities are selected based on the size of the expression and evenly distributed across three sets 1) single symbols, 2) small formulas and 3) large formulas.

Single symbols are collected by a method of rejection sampling from the set of symbols extracted from Wikipedia page[1]. Formulas are selected only from the Wikidata source as it is the largest sample of formulas and the

[1] https://en.wikipedia.org/wiki/List_of_mathematical_symbols

extraction process is more robust than for Wikitionary and Proof Wiki. Since its difficult to classify the length of an expression based on the LaTeX representation, we convert each expression into its Symbol Layout Tree (SLT) representation [6] and then determine the number of nodes in the SLT.

Table 5.1 shows the descriptive statistics of the formulas based on the number of nodes in the SLT. We consider the 33.33rd and 66.66th percentile of all the formulas extracted from Wikidata, to distribute the data into three sections. Small formulas (between 2 and 10 nodes), medium (between 10 and 20 nodes), large (20 and above nodes). Since the SLT representation considers every fraction, function or variable as an individual node, we select only formulas from the small and large sets so that the differences in sizes are visually significant. Take for example Equation 5.1 which contains 10 individual nodes in the SLT but compared to Equation 5.2 which has 17 nodes, the visual differences are not easily observable. However the differences between Equation 5.1 and Equation 5.3 is visually significant. We hence exclude any formulas from the medium section and only consider formulas from the small and large section.

$$u_{xx} + xu_{yy} = 0 \tag{5.1}$$

$$G = \pi_1(X)/p_*(\pi_1(C)) \tag{5.2}$$

$$\forall (x, y, z) \in X^2 \times Y, \quad x\Re z \wedge y\Re z \implies x = y \tag{5.3}$$

Table 5.1: Descriptive Statistics of formulas sizes (without single symbols) in the Wikidata Data set

Min. Value	**Max. Value**	**Mean**	**Variance**	**Standard Deviation**
2	264	18.27	175.86	13.26

Further, each set of symbols, small formulas and large formulas are equally divided into two halves, familiar concept/formula & less familiar concept; this is done to observe if there is an effect based on familiarity of a concept. We classify familiarity based on whether a mathematical concept would be encountered during years 1 or 2 of a standard college education. We assume, if a mathematical concept is familiar, it is familiar across both concept name and the formula associated with it. Since this might not always be the case, we measure participant's responses across three levels, "I've never seen it before", "I've seen it before but I am not sure of its meaning", "I've seen it before and know its meaning". The classification of symbols, small formulas, and large formulas with each having a familiar and less familiar category was maintained across the practice trials as well.

The experiment and data collection was performed on an online web interface, designed and developed by the experimenter. The system was developed with Python-Flask, SQLite, HTML and Bootstrap. The computer connected to the monitor was running Windows 10, and the participants took the survey on a Firefox Browser with a standard keyboard and mouse. Materials used outside the system were the consent form, a sign-off sheet to track payments need as per university financial policies and a copy of the Thank you page with contact information that was handed out after participants had completed the study.

5.1.1 Mathematical Entity Selection

A Latin-square design (Table 5.2) is used to balance the styles in which a card type for a mathematical entity is presented to a participant. Each value in a single row, sequentially contains the card type to be displayed for its corresponding entity (entity ID present in the column header). This ensures a balanced presentation style across both participants and entities. We have have 6 card types across 2 query input types (LaTeX and text) resulting in 12 card types overall but ID's 7-12 are a repeat of 1-6. Due to the limited number of card types a participant might be able to figure patterns in presentation order if presented sequentially. We hence randomly shuffle each row before presenting cards to participants, to minimize any bias introduced due to card type ordering.

5.2 Participants

Participants were recruited via emails sent out to both students and faculty within the College of Science and Golisano College of Computing and Information Sciences at Rochester Institute of Technology. There was no prescreening done, since we wish to see information preference levels for math information retrieval of participants irrespective of number of math courses taken. Participants were scheduled to take the experiment one at a time within

Table 5.2: Counterbalanced order used to present card types to participants (P) for corresponding mathematical entities (E)

	E1	**E2**	**E3**	**E4**	**E5**	**E6**	**E7**	**E8**	**E9**	**E10**	**E11**	**E12**	...	**E47**	**E48**
P1	1	2	3	4	5	6	7	8	9	10	11	12	...	11	12
P2	2	3	4	5	6	7	8	9	10	11	12	1	...	12	1
P3	3	4	5	6	7	8	9	10	11	12	1	2	...	1	2
P4	4	5	6	7	8	9	10	11	12	1	2	3	...	2	3
P5	5	6	7	8	9	10	11	12	1	2	3	4	...	3	4
P6	6	7	8	9	10	11	12	1	2	3	4	5	...	4	5
P7	7	8	9	10	11	12	1	2	3	4	5	6	...	5	6
P8	8	9	10	11	12	1	2	3	4	5	6	7	...	6	7
P9	9	10	11	12	1	2	3	4	5	6	7	8	...	7	8
P10	10	11	12	1	2	3	4	5	6	7	8	9	...	8	9
P11	11	12	1	2	3	4	5	6	7	8	9	10	...	9	10
P12	12	1	2	3	4	5	6	7	8	9	10	11	...	10	11

a 30 minute time slot. Scheduling of participants was done with the help of an online scheduling software Doodle[2]. Each participant was compensated $10.00 for their participation in the study. Appendix A and Appendix B contain the email and poster respectively, used to recruit participants.

5.3 Variables & Confounds

The six card types across two query types (text vs. LaTeX) along with the two levels of familiarity (familiar vs less familiar) and three levels of formula size (symbols, small formulas and large formulas) were the controlled or independent variables (IV). Usefulness value of a card, with four levels measured on a Likert Scale between 1 and 4 (1 being not useful, 4 being highly useful), content understanding with two levels (yes or no), and time to respond

[2] https://doodle.com/

to the queries are the measured or dependent variables (DV).

The cards were designed in a neutral manner (without any color) to remove any confounds, arising due to font size, individual section boxes for title, formula & description. The contents are placed in fixed size boxes without making the addition of any new component obvious to a participant. That is, we do not add borders for individual components as shown in the design chapter. We use pre-generated images of the LaTeX query, to avoid having a conversion delay due to rendering which might expose the $\LaTeX$ input to a participant. All cards used in the experiment are included in Appendix D.

5.4 Procedure

Participants were scheduled to meet one-on-one with the experimenter during predetermined time slots: between 9:00am and 12:00pm or 4:00pm and 7:00pm. The meeting took place over 7 days with up-to 7 sessions per day. The experiment was conducted in the Computer Science Break Out Rooms in the Golisano College of Computing and Information Science building at RIT. Once there, the participants were instructed to take a seat in front of a Monitor connected to a laptop for the experiment. Participants were then introduced to the experiment, informed about the anonymity of their participation, the expected duration of the experiment, and the compensation process. They were then given the consent form for them to read and provide consent. All through this time the experimenter answered any questions the participants had regarding the experiment or the process.

The experimenter verbally reminded the participants that the evaluation is purely of the system under test and is in no way intended to serve as a test of their mathematical knowledge. The participants were also encouraged to take their time to carefully consider each scenario before responding but to respond as quickly as possible as it is a timed task. This reminder along with the instructions were present on the landing page the participants see, before filling out the demographic survey. Participants were then briefed about the order of the experiment, in terms of seeing the Demographic survey, followed by four practice trials to help them familiarize themselves with the interface, followed by the experimental trials, at the end of which was a post-study questionnaire. Participants were also informed that no questions could be answered after the practice trials were done. All of this was part of a pre-written script to ensure that all participants receive the same information and in the same order (see Appendix C).

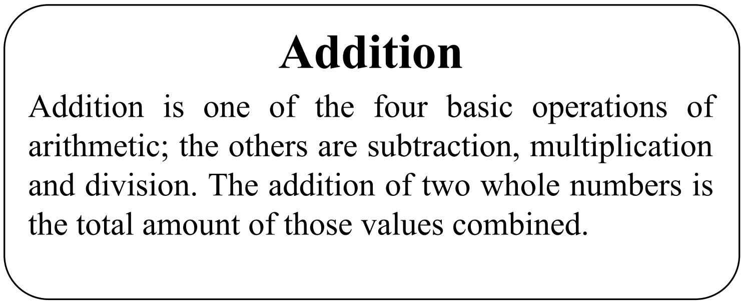
5.5 Trials

Every trial would begin by showing a query for a mathematical entity e.g. for addition a text query would be 'Query: Addition' and LaTeX would be 'Query: $+$'. Next a participant would respond to the question :

- What is your level of familiarity with this concept?
  - I've never seen it before.

  - I've seen it before, but I'm not sure of its meaning.
  - I've seen it before and know its meaning.

Participants would then have to click on 'Next Section' to proceed, time is recorded till 'Next Section' is clicked to analyze how quickly participants respond to text vs LaTeX queries. The next section displays a single card type as shown for addition in Figure 5.2. Participants were then asked to evaluate the card and provide responses to the three questions:



Addition

Addition is one of the four basic operations of arithmetic; the others are subtraction, multiplication and division. The addition of two whole numbers is the total amount of those values combined.

Figure 5.2: Card for Addition containing title and description

- How useful is this card in providing information about the query?
  - Not useful
  - Slightly
  - Moderately
  - Highly
- Is the information on this card understandable?

  - No
  - Yes

- (Optional) Do you have any additional comments about this card?

A participant then had to click on ‘Next Question’, which would record the time for this section. A question counter was present in the lower half to help keep track of the current and total questions. Since typing speeds vary across individuals, we understand there could be an difference in response time due to the comments, and left it optional.

5.6 Post-Study Questionnaire

The post-study questionnaire consisted of two main sections in the first the participants were asked to rate (on Likert Scales) the importance of the presence of title, formula, and description on the card; this is done to observe the overall effect as perceived by the participant. The second section asked about the usefulness of having links to related concepts, links to resources such as tutorials, proofs and other resources, and more formal mathematical descriptions along with existing mathematical descriptions on the card. This was done to consider possible future directions of research. Examples of the questions are present in Table 5.3 & Table 5.4.

Table 5.3: Questions from Section 1 of Post Study Questionnaire.

	Not Important	**Slightly Important**	**Moderately Important**	**Important**	**Very Important**
Title on a card					
Formula on a card					
Description on a card					

Table 5.4: Questions from Section 2 of Post Study Questionnaire.

	Not Useful	**Slightly Useful**	**Moderately Useful**	**Very Useful**
Links to **related concepts**				
Links to **resources such as tutorials, proofs**				
Formal (mathematical) descriptions				

5.7 Summary

In this chapter we described the protocol we followed for the human experiment design. We also explained our design choices for independent and dependent variables, selection of mathematical entities and overall question selection in both experiment and post study questionnaire. In the next chapter we discuss our results and observations across participants for usefulness of card components.

Chapter 6

Results

In the following sections we present the results obtained from our human experiment and use statistical test to validate the findings were due to the independent variables and not due to participant variances. We conclude with results from the post study questionnaire and discussions of the results.

6.1 Demographics

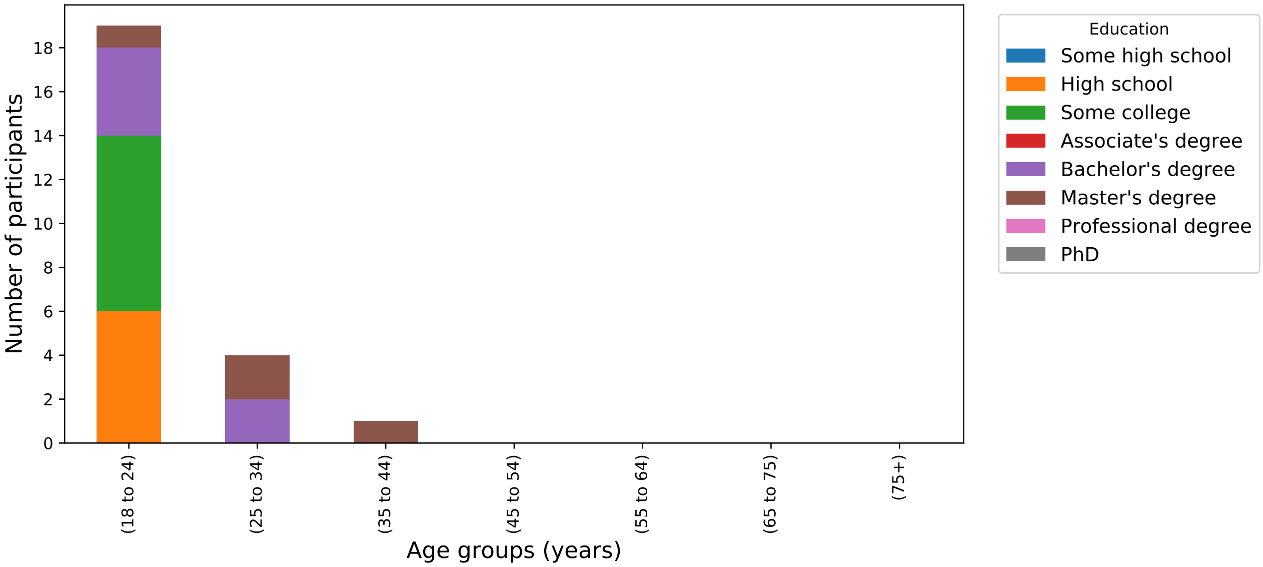


Figure 6.1: Age & education of participants

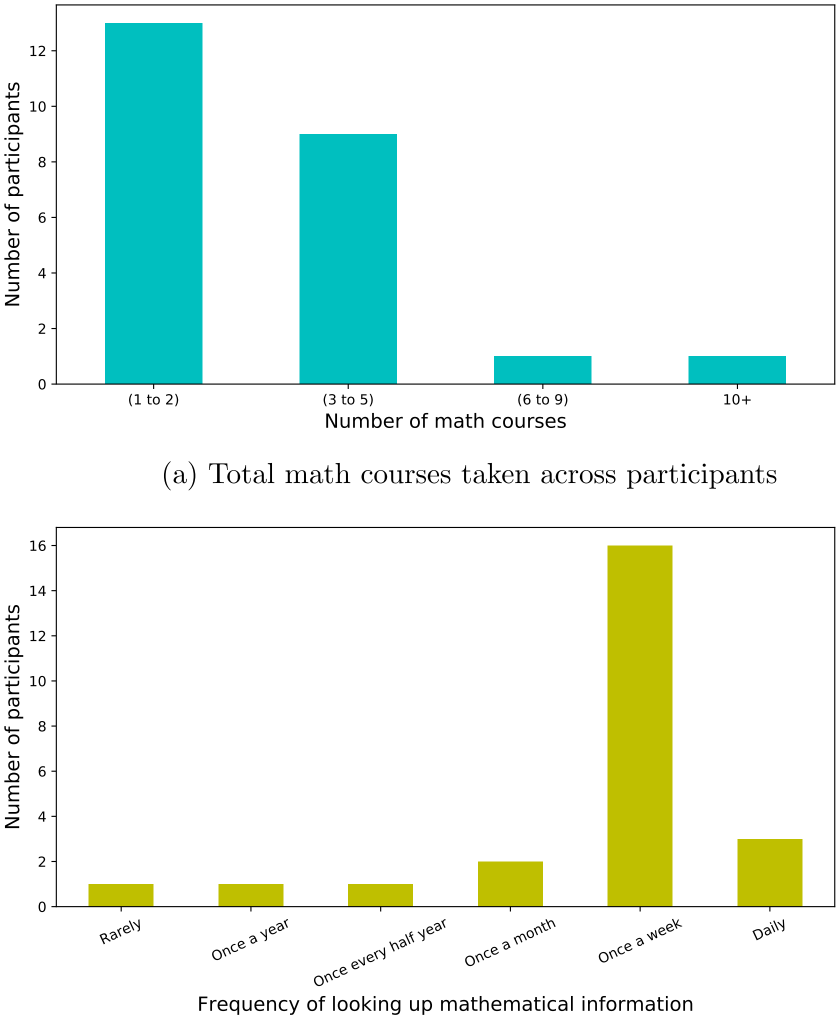
A total of 24 participants completed the experiment. 58.33% (n=14) of the participants reported their gender as male, 37.5% (n=9) of the participants reported their gender as female and 4.16% (n=1) of the participants reported their gender as other (non-binary).

79.16% (n=19) of the participants reported being between the ages of 18 and 24, 16.66% (n=4) of the participants reported being between the ages of 25 and 34 and 4.16% (n=1) of the participants reported their age to be between 35-44.

25% (n=6) of the participants reported to have completed High School and are Freshmen, 33.33% (n=8) of the participants reported to have completed Some College, 25% (n=6) of the participants reported to have completed a Bachelor's Degree and 16.66% (n=4) of the participants reported to have completed a Master's Degree. See Figure 6.1 for more details on the distribution of age and education.

Figure 6.2a shows the distribution of the number of participants and the number of math courses taken, about 50% (n=12) of the participants have taken at-least 1 to 2 math courses, the rest have taken more than 2. This is useful to know since we control for familiarity of math concepts and wish to observe the effect of math entity cards on both familiar and less familiar concepts.

Figure 6.2b shows 66.66% (n=16) of the population look up mathematical information at least once a week. Participants were provided with the following examples of mathematical information as part of the demographic



(a) Total math courses taken across participants

(b) Frequency with which participants need to look up mathematical information

Figure 6.2: Bar plot of math courses taken and frequency of looking up mathematical information as reported by participants

survey, function definitions (e.g. trigonometric and statistical functions), definitions for mathematical symbols, function plots, mathematical models (e.g environmental or physical models), theorems, and proofs. Only three participants felt that they would look up math information less than a month (Once every half year, Once a year and Rarely).

Figure 6.3 shows the frequency response of the participants to the question 'How frequently do you need to express mathematical notation when using a computer, such as for writing technical documents or in using computer programs such as Matlab, Mathematica or Maple?'. 75% (n=18) feel the need to express mathematical notation when using a computer at least once a month if not more (once a week, daily). Thus demonstrating the usefulness of having a math aware search engine, which would make looking and expressing mathematical notation simpler and faster for these participants.

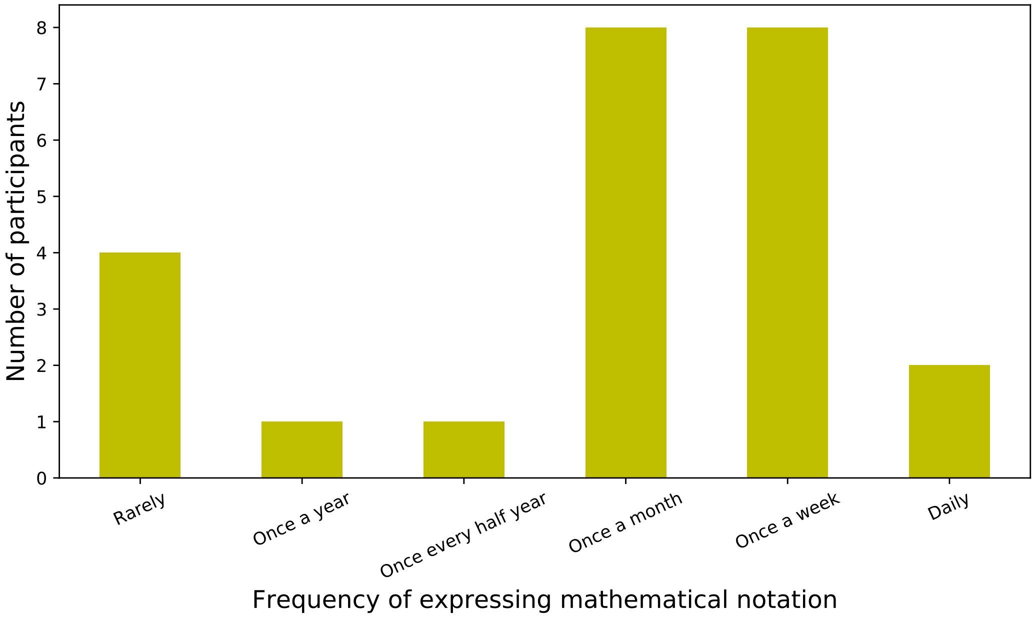
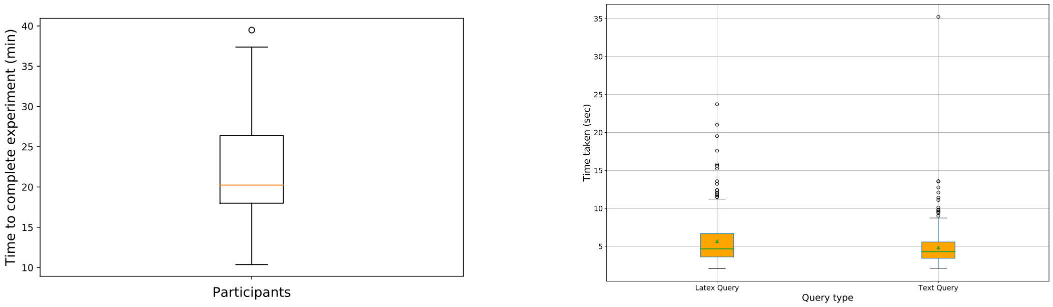


Figure 6.3: Frequency with which participants need to express mathematical notation

6.2 Experiment

With regards to the previous chapter on Human Experiment, this section summarizes our observations across all the independent variables and how useful participants find individual card components. We measure both individual response times to a query and overall time duration to complete the experiment. Individual query times were measured independently for each section. Section 1 of a trial asks whether participants are familiar with the concept/formula and section 2 measures the usefulness of a card type. This helps us compare time differences between recognizing text and LaTeX queries.



(a) Distribution of overall time to complete the experiment

(b) Time to recognize a LaTeX vs text query type as familiar

Figure 6.4: Box plot of overall and indiviual time taken for section 1 of each trial

As seen in Figure 6.4a, a majority of the participants completed the experiment in less than 30 minutes. The difference between the shortest and longest time taken to complete the experiment can be attributed mainly to participants providing comments in each trial to the optional section ‘Do you have any additional comments about this card?’.

Next we wish to observe the difference in times for participants in interpreting a text query vs a formula query as familiar. Since communicating with text is more popular than formulas (LaTeX) we expect people to recognize text queries more quickly. We classify the response "I've never seen it before" as a participant being less familiar and responses "I've seen it before, but I'm not sure of its meaning" or "I've seen it before and know its meaning" as a participant being familiar with the query. As seen in Figure 6.4b the time taken to recognize a query as familiar is overall slightly larger for a LaTeX query than for a text query, supporting our initial assumption.

From the 48 concepts in total (refer to Appendix E), we have three sets of 16 across Symbols, Small Formulas and Large Formulas. Each set of Symbols, Small Formulas and Large Formulas is equally divided (50-50) into familiar and less familiar concepts. However not all concepts classified by us as familiar would necessarily be familiar to a participant. It would depend on their exposure to the concepts and their formulas as well. We analyze this difference to find 21.70% (n=125) queries are found to be not familiar (response = "I've never seen it before") for 576 (48 × 24) queries we classified as familiar and 8.5% (n=49) queries are found to be familiar (response ="I've seen it before and know its meaning") for 576 queries we classified as less familiar. The relatively low percentage of both (21.70% and 8.5%) might affect our analysis in a minor way but is a factor that is hard to control. For further experiments we might use either a more strict criteria for selecting familiar and unfamiliar concepts or we would filter participants beforehand.

Due to this difference which is spread across participants we measure familiarity and less familiarity based on our classification which is a 50-50 distribution across each set for all further results and observations.

6.2.1 Usefulness of Card Components

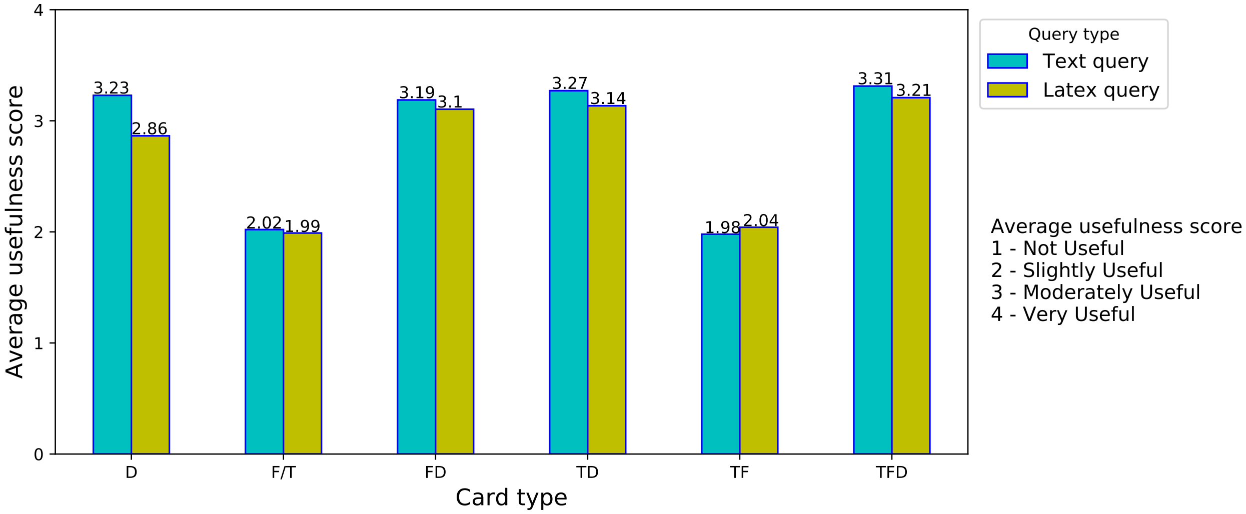


Figure 6.5: Overall Usefulness Scores per Card Type

In this section we analyze the usefulness of each component of a card title(T), formula (F) and description (D) as well as all combinations of title-formula (TF), title-description (TD), formula description (FD) and title-formula-description (TFD). We compare this across both query types and familiarity levels as well as across different formula sizes. As seen in figure 6.5 the description has the highest usefulness value (3.23) as compared to just the title (1.99) or the formula (2.02). The difference between description usefulness scores across query types could be attributed to participants expecting descriptions of formulas to explain not just the mathematical entity but also the variables and the relationship between them in the formula. On the other hand for a text query, participants are mainly concerned with a description that tells them "something" about the mathematical entity. Receiving all three title, formula & description is valued the highest across card types which is similar to our assumptions.

6.2.2 Understanding of Content

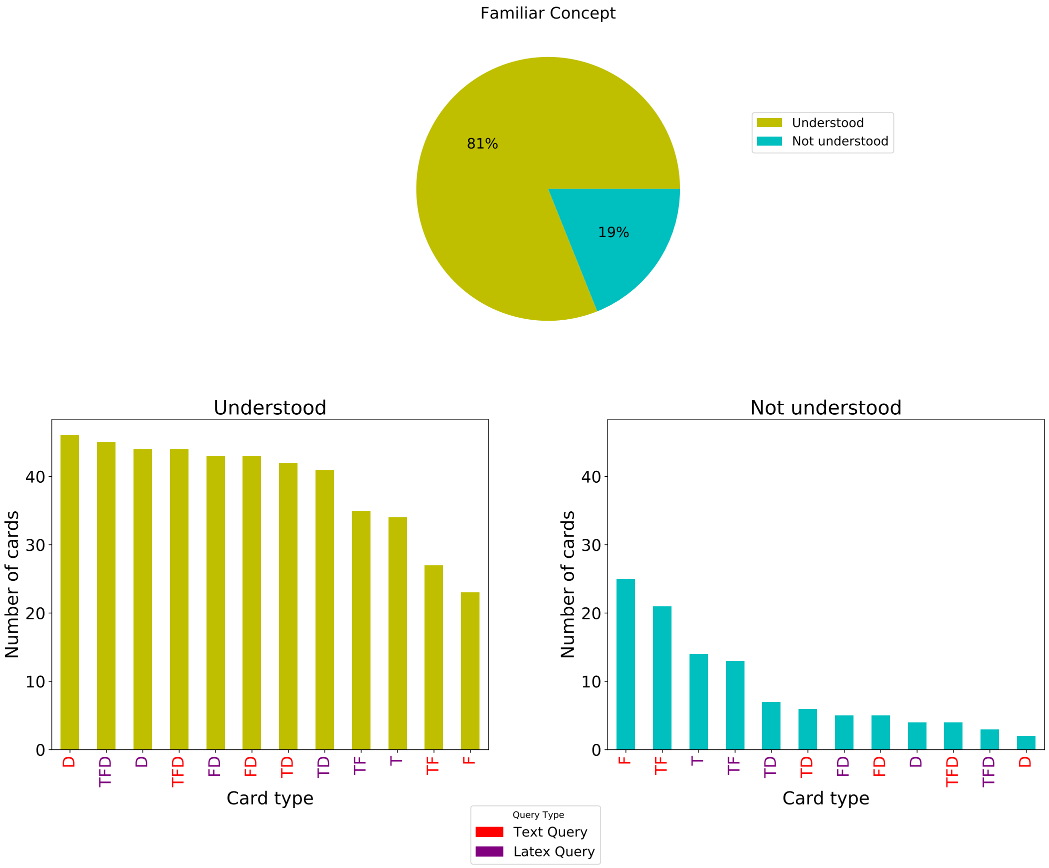


Figure 6.6: Card-types contribution to understanding, for familiar concepts

We also measured responses to a question “Is the card understandable” with a binary response option of Yes or No. This was done to check whether there is a difference in card types in understanding content, we suspect cards containing only formulas to received the highest number of ‘No’ responses, since a formula is usually ambiguous without its surrounding text. We analyze this response further with respect to familiar and less familiar concepts.

The pie plot in Figure 6.6 shows how many of the queries classified by us as Familiar, were understood by the participants. The bottom histogram shows the distribution of each card types to understanding and not understanding. As we see having the description for a text query contributed the most to understanding, but having the title, formula and description contributes the most for a LaTeX query, closely followed by having just the description. For a text query having a formula without and with a title, contributes the least to understanding a concept and also contributes the most to not understanding a concept. This could mean that overall, for understanding content a formula should always preferably be accompanied by some text description. Having both the formula and description for both text and LaTeX contribute equally (4.59%) to not understanding, this we presume to be the case when the description does not explain the symbols in the formula but just the concept.

We plot an analogous plot for the less familiar concepts as well, to check for any differences. With reference to Figure 6.7 having the Formula and Description for a LaTeX query and the analogous Title and Description for a text query (the first two bars) contribute equally to understanding. Thus re-

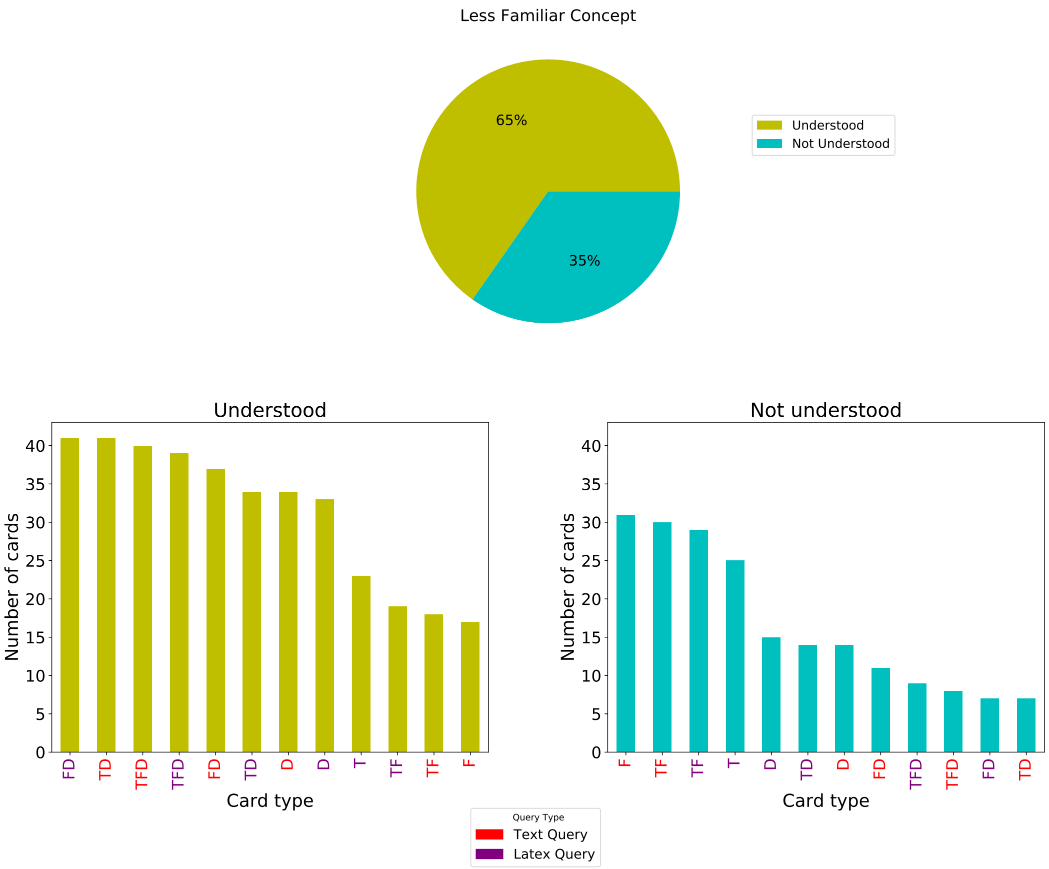


Figure 6.7: Card-types contribution to understanding, for less familiar concepts

confirming the importance of a description in both query types. This is closely followed by having title, formula and description for a text query. Similar to Familiar concepts having the formula with or without the title contribute the least to understanding and the most to not understanding a concept.

Overall we see some common explainable patterns, having the description helps understanding but however if the description is incomplete in terms of missing variable names and interactions between them, this causes a break in the understanding of the content, and possibly opens up more questions for a participant. Having multiple descriptions could be beneficial but more importantly it would be beneficial to have explanations for variables names in the description itself.

6.2.3 Participant Comments

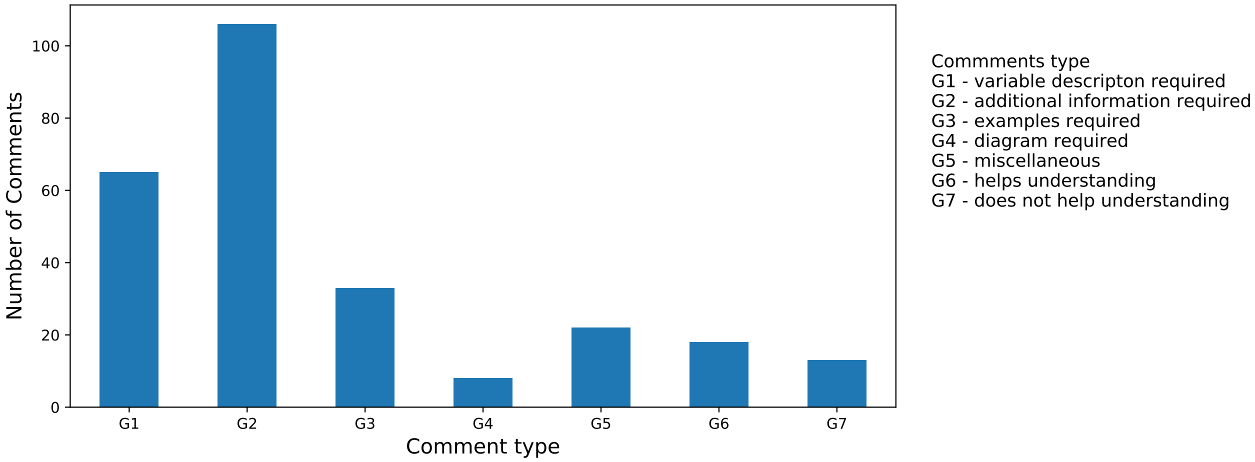


Figure 6.8: Comment Distribution across Groups

In this section we provide our analysis of the comments provided by participants for individual card types. Overall we received 200 comments for a total of 1152 queries. To simplify the analysis we categorize comments into one of the following 7 groups:

1. Variable description required
2. Additional information required
3. Examples required
4. Diagram required

5. Miscellaneous
6. Helps understanding
7. Does not help understanding

Figure 6.8 shows the distribution of comments per group. Overall the comments for additional information required is almost 50% (n=110). 32.5% (n=65) of the queries ask for explanation of variables for formulas. About 15% (n=30) of the comments suggest adding examples to existing descriptions.

To understand comment distribution per card type refer to Figure 6.9. For a text query, receiving the title and formula (card type 2), causes participants to explicitly ask for an additional explanation. This we assume is because we do not provide them with any description. We also see a higher number of comments asking for an explanation of variables, which supports our assumption. For text query when only a formula is returned (card type 5) participants ask for additional explanation, variable description as well as examples of the formula. This is logical since a formula without any text is ambiguous and requires some explanation to help understand the formula better.

For a LaTeX query when title-formula-description (Card Type 7) is provided participants ask for variable information to be provided, this observation is higher as compared to text query indicating a difference in what is being expected as query type changes. For a LaTeX query when title-formula (card type 8) is provided and only title (card type 10) is provided, participants ask for additional information

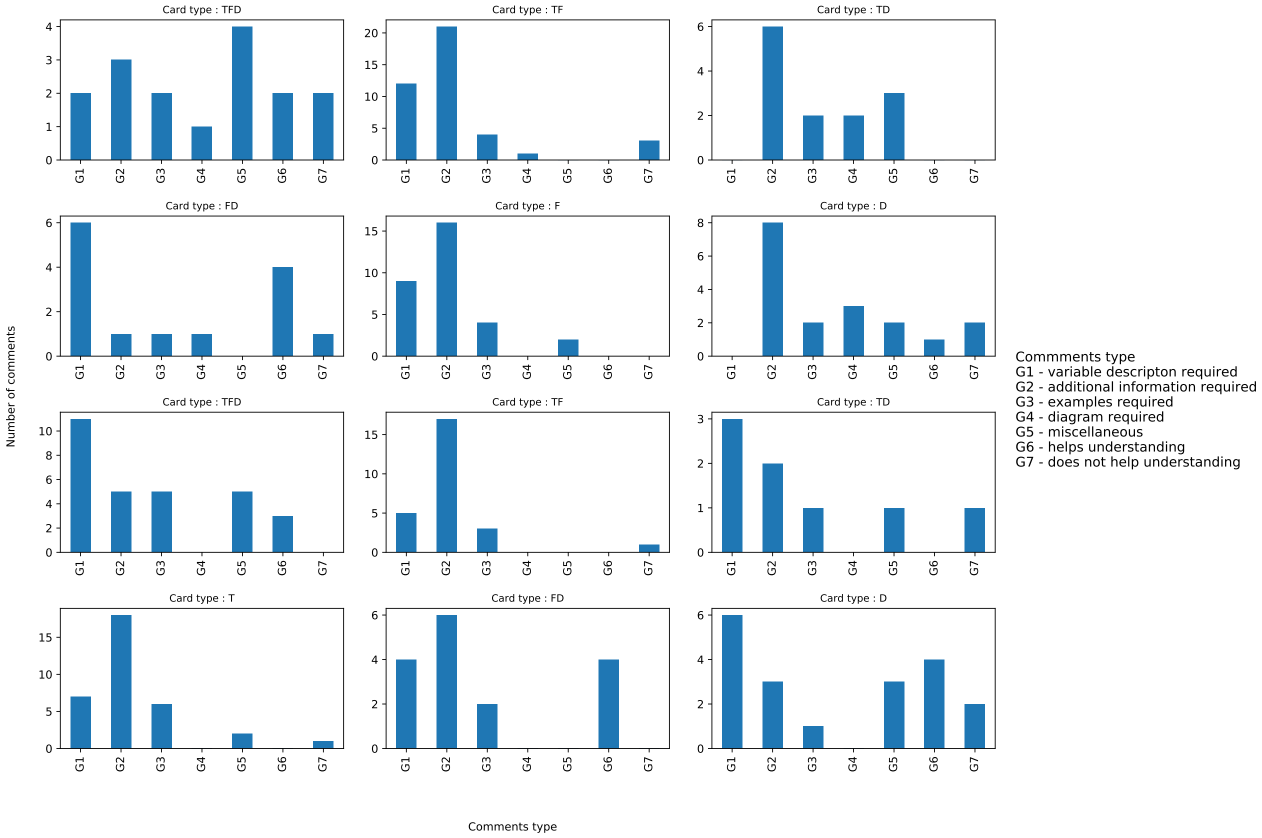


Figure 6.9: Comment Distribution across Card Types

6.2.4 Secondary Results

In this section we summarize our findings of the distribution of usefulness scores, across familiarity and formula sizes for both query types, text and LaTeX. Appendix F contains the plots of distribution which will be referred in this section.

6.2.4.1 Between Familiar & Less Familiar Concepts

From Figure F.1 and F.2, we see a sharp decrease in very usefulness score (density of dark red), indicating that the same card types are affected based on prior familiarity of the mathematical entity. One possible assumption for this could be for familiar concepts the card contents help refresh an prior understanding, where as for less familiar concepts, a participant is trying to understand the content. As seen in Figure F.3 and F.4 card types with description are rated more useful than card types without description.

6.2.4.2 Between Symbols, Small and Large Formulas

With reference to Figures F.5, F.6 and F.7 for a LaTeX query the very usefulness score (density of dark red) decreases from symbols to small formula and from symbols to large formula for card types that contain the description. One possible explanation for this could be, as the number of variables in a formula increases, the description must explain every variable contained.

6.3 Statistical Testing

To see the impact of card types and query types on usefulness and response times, we conduct a two-way repeated measures ANOVA to verify that the difference in usefulness scores and response times, is due to the independent variables, and not due to inter-participant variation.

6.3.1 Usefulness score

The two-way repeated measures ANOVA, shows significant evidence against the null hypothesis H_0: card types or query types has no impact in usefulness scores. We find both query (F(1; 23) = 4.63, p=0.0042) and card type (F(5; 115) = 62.87, p=2.71e-31) to have an effect. However there is not a significant impact due to any interaction between query and card types (F(5; 115) = 2.00, p=0.08).

We conduct Wilcoxon Signed Rank test as a post-hoc, and receive a p-value=0.007, which shows that the median usefulness score for text query is greater than the median usefulness score for a LaTeX query. This is against our initial assumption that math entity cards are more useful for math information retrieval where in search revolves around formulas (LaTeX).

To check the impact of individual card types on usefulness score, we conduct a Pairwise Wilcoxon Signed Rank Test with Bonferroni correction, as our variable (cardtype) has multiple levels (6). As we see in Table 6.1 every card component is compared to every other card component. Note: For the statistical test we do consider receiving only the formula for a text query and receiving only the title for a LaTeX query to be in the same group.

Table 6.1 shows significant evidence to reject the null hypothesis : there is no difference in the median usefulness score in card types. Table 6.2 shows the card type pairs for which the test showed significant and no significant difference. From the two tailed Pairwise Wilcoxon Signed Rank Test with Bonferroni Correction, we found which card types have difference in the me-

Table 6.1: Pairwise Wilcoxon Signed Rank Test with Bonferroni Correction for usefulness scores

	TFD	**TF**	**TD**	**FD**	**F/T**
TF	0.00039	-	-	-	-
TD	1.0000	0.00039	-	-	-
FD	1.0000	0.00027	1.0000	-	-
F/T	0.00040	1.0000	0.00033	0.00040	-
D	0.65541	0.00045	0.83819	1.0000	0.00026

Table 6.2: Observations : Pairwise Wilcoxon Signed Rank Test with Bonferroni Correction

Significant difference	**No significant difference**
TF - TFD	TD - TFD
TD - TF	FD - TFD
FD- TF	FD - TD
F/T - TFD	F/T - TF
F/T - TD	D - TFD
F/T - FD	D - TD
D - TF	D - FD
D - F/T	

dian. We then performed a right tailed Pairwise Wilcoxon Signed Rank Test with Bonferroni Correction and for some card types, found significant evidence to reject the null hypothesis H_0: difference in the median usefulness score between card pairs is less than 0, the conclusions are shown as follows :

$$\left.\begin{array}{l} TFD \\ TD \\ FD \\ D \end{array}\right\rangle \begin{array}{l} TF \\ F/T \end{array}$$

Thus indicating that overall a description increases the usefulness score significantly.

6.3.2 Response Times

The two-way repeated measures ANOVA, shows significant evidence against the null hypothesis H_0: card types or query types has no impact on response times. We find only card types (F(5; 115) = 9.04, p=2.856e-07) to have an impact on repsonse times. There is no evidence of an effect of query type (F(1; 23) = 0.006, p=0.9405) or any effect due to interaction between query and card type (F(5; 115) = 1.452, p=0.211) on response times for usefulness evaluation.

To check the impact of individual card types on response times, we conduct a Pairwise T-test with Bonferroni correction, as our variable (card-type) has multiple levels (6). As we see in Table 6.3 every card component is compared to every other card component. Note: For the statistical test we do consider receiving only the formula for a text query and receiving only the title for a LaTeX query to be in the same group.

Table 6.3: Pairwise T-Test with Bonferroni Correction for response times

	TFD	**TF**	**TD**	**FD**	**F/T**
TF	5.1e-05	-	-	-	-
TD	1.0000	0.0384	-	-	-
FD	1.0000	0.0032	1.0000	-	-
F/T	3.5e-05	1.0000	0.0210	0.0228	-
D	1.0000	0.0016	1.0000	1.0000	0.00025

Table 6.3 shows significant evidence to reject the null hypothesis: no difference in the mean response times due to card types. Table 6.4 shows the card type pairs for which the test showed significant and no significant difference.

Table 6.4: Observations : Pairwise Wilcoxon Signed Rank Test with Bonferroni Correction in Response times

Significant difference	**No significant difference**
TF - TFD	TD - TFD
TD - TF	FD - TFD
FD - TF	FD - TD
F/T - TFD	F/T - TF
F/T - TD	D - TFD
F/T - FD	D - TD
D - TF	D - FD
D - F/T	

From the two tailed Pairwise T-Test with Bonferroni Correction, we found which card types have difference in the mean. We then performed a right tailed Pairwise T-Test with Bonferroni Correction and for some cards found significant evidence to reject the null hypothesis H_0: difference in the mean response times between card pairs is less than 0, the conclusions are shown as follows :

$$\left.\begin{array}{l} TFD \\ TD \\ FD \\ D \end{array}\right\rangle \begin{array}{l} TF \\ F/T \end{array}$$

Thus indicating that overall having a description increases the mean response times in evaluating card usefulness significantly.

6.4 Post Study Questionnaire

This section summarizes the overall importance given to each section (Title, Formula and Description) by participants after completing the experiment. It also summarizes the usefulness of having additional features specifically : links to related concepts, links to other resources such as tutorials and more formal descriptions on the card itself, we believe this would help focus future efforts on improvising the cards.

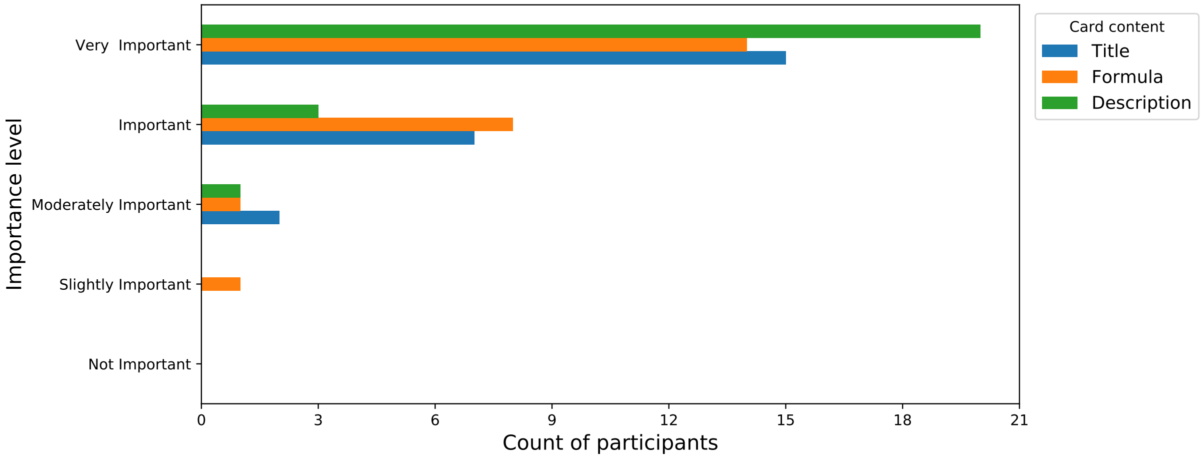


Figure 6.10: Importance of title, formula and description

As we see in Figure 6.10 the most important feature overall, is the description of the concept. According to the participants, the title is the second most important feature on a card. We compute the average score (score × number of participant/Total number of Participants) for each, Title=4.54, Formula=4.46, Description=4.79. We find the difference between title-formula (0.08) is small compared to description-title (0.25) or description-formula (0.33). This could mean having the formula is almost as important as having a title on the card for a mathematical concept, as the formula helps provide an additional attribute/property for the concept, in a manner similar to the title.

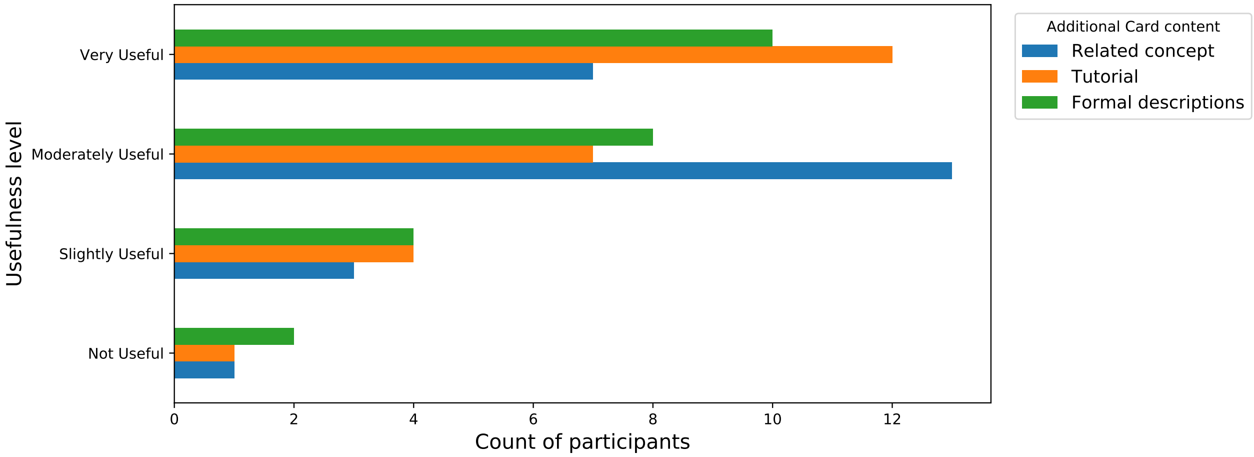


Figure 6.11: Usefulness of related concepts, tutorials and formal description

Figure 6.11, shows us how useful having additional features on the card would be. We again compute the average score to make the comparison easier, Related Concept=3.08, Tutorial=3.25, Formal Descriptions=3.08. It is interesting to note overall participants find having formal descriptions for the same concept equally useful as compared to having other related concepts on the card. The score for tutorial is higher than the other two features but this is no surprise as the sample population is mainly students, who would be looking for other resources to understand the concept further.

6.4.1 Participant Comments

This section discusses the additional comments provided by some users. Six comments were provided in total and they are as follows :

1. Potentially less formal descriptions in some cases, maybe a setting with

3 levels?

2. Examples of the equation.

3. I personally also find it very useful to have explanations of symbols involved in equations.

4. Examples wherever possible would be highly useful too.

5. It would be cool if the system could tailor the results to my level of understanding of concepts, so that it could provide better explanations for concepts/areas that I'm very unfamiliar with (while not cluttering things that I am familiar with, where I'm probably just looking to remind myself of the formula).

6. Having simple diagrams to help visualize the different terms and variables might improve clarity, concept understanding, and recognition in a lot of cases.

Point 3 is similar to the comments provided during the experiment where participants are also interested in understanding the explanation of what symbols in the equation mean or represent. Points 2 and 4 suggest examples, which is similar to the usage section of the card as discussed in the Design Chapter. Interestingly Point 1 and 5 talk about less formal descriptions, and adaptations to less familiar concepts. Although all the data for the experiment, was obtained only from Wikipedia, it suggests sometimes it is also helpful to have a more simpler version of the description. This supports the idea of using

Wiktionary as an alternate data source. It also brings about a point of consideration that users are not always looking for more technical resources which would suggest indexing more than the latest scientific articles.

As discussed earlier we do agree with Point 6, regarding the use of diagrams and images but feel in general some concepts can be represented by a diagram easily, this opens up the opportunity of tailoring the resources provided based on the field of mathematics, Geometry for example would be an ideal candidate for a diagram to supplement understanding. Linear Algebra and Probability would possibly be better explained by an example instead of a diagram in some cases.

6.5 Discussion

In the design chapter, we assumed under a math informational retrieval setting math entity Cards would help address a factual informational need. We started out with three basic card designs that can basically be expressed as receiving titles with formulas, receiving title-formula-descriptions and receiving a usage section in addition to the title-formula-description. We believe that there are questions regarding the usefulness of this alternate design (with formulas added) in general and hence for the human experiment only consider combinations amongst the first two types (title-formula & title-formula-description). In this section we discuss our overall observations across our designs, and summarize what we feel to be the most promising card design as well as ideas/directions for future research.

Our initial assumption was a title-formula card might suffice in some cases. For the case of familiar concepts we expected a moderate amount of usefulness as the title serves as a form of confirmation for a LaTeX query. However as seen across all the charts on usefulness scores, the number of participants that find it useful (very useful and moderately useful) are less than 50% across both query types text and LaTeX. This suggests that its rarely useful and preferably avoidable to present just the title, just the formula or title & formula only. This would be the situation had we relied only Wikidata as a data-source for math entity cards, since all of Wikidata formulas have titles but a majority of them lack descriptions.

For the second case of title-formula-description we decided to extract content from Wikipedia as it is one of the most cited open sources of general content. We also fetched descriptions for concepts from Wiktionary and Proof-Wiki with the assumption that having multiple levels of descriptions will be useful to users. The carousel and pop-up modal feature were described with which a user could consume multiple different descriptions for a single concept. However for the experiment, we decided to control the description section by having only the lead section of Wikipedia used. Due to the relatively low popularity of math information retrieval in general, we decide its better to add features that are valid and applicable rather than just adding features that we hope are useful.

We assumed when looking for formulas associated with mathematical concepts, a general description of the concept would suffice. We were partially correct with this assumption. As seen by the description section receiving the most usefulness score both individually and when combined with either title or formula. However interestingly we also found that for LaTeX queries, having a general description of the concept does not suffice. Participants do also require an explanation what the variables are, the relationship between them and their units of measurements if any.

From the comments section we noticed a few participants were not able to understand even Wikipedia descriptions. This is means we might have to use additional educational resources such as open textbooks that are used to explain topics to students. However on the flip side, a learned participant might not find having Wikipedia descriptions useful at all, this needs to be validated by conducting further experiments.

Since our human experiment only considers description and not definitions, for future work we suggest a direct comparison of usefulness between descriptions of concepts and definitions of formulas. Definitions are relatively more concise and detailed as compared to general descriptions. Most definitions of formulas also describe the variables used within the formula, which why we feel there might be some interesting findings. We would also recommend a comparison between multiple descriptions for familiar concepts vs multiple descriptions for less familiar concepts. This would help highlight under what circumstances is multiple descriptions actually useful. Do users want them all the time or when they are looking up something they do not know anything about?

Participant comments provide us both positive and critical feedback. Some participants notice when a title is missing, or when a formula is missing, but most of the critical feedback occurs when the description is missing. Overall we find having examples and links to existing less commonly known terms would be very useful. We did suggest a section on 'Usage' within the design of cards, that could account for examples but feel instead of providing links to less commonly known terms within the card, it could also be better to combine them with related concepts in an alternate tab.

6.6 Summary

Overall we notice having the description to be the most useful in all card components across query types. We also find statistically significant evidence which suggest that card types for LaTeX queries are considered to be more useful than cards received for text queries. There is also statistically significant evidence that card types affects response times, however an increase in response times is not always a negative outcome, especially if the information need is satisfied.

Receiving the description is a logical assumption in information retrieval, but when searching for a formula participants would also prefer the description to explain the variables in the formula. Further studies would be required to confirm if a participant prefers a description of the formula over a description of a concept. Given our observations, it would make sense to have an equal distribution of both formal or technical and relatively less formal descriptions as well. In the next chapter we conclude and put forward our suggestions for improving math entity cards.

Chapter 7

Conclusion and Future Work

We introduced an alternate design for entity cards describing Mathematical Concepts. We believe the new design is better suited as compared to the regular design, for the field of mathematics, as formulas are central to creating and communicating information within the field. The new card design helps ease the transition between formulas and their associated text (titles or descriptions).

We demonstrate the creation of these cards using knowledge bases in structured and semi-structured format. Due to the inherent complexity in understanding mathematics, we resort to extracting multiple descriptions for the same concept, when possible. With the help of language and context cues such as contents within bold tags and keywords such as 'defined', 'denoted by', we disambiguate and select a single formula that best represents the associated concept. Thus allows multiple formulas to be indexed and be associated with a single concept. These multiple descriptions, with different degrees of formality are intended to support the information needs of both beginner and intermediate users.

In the context of a math-aware search engine, where search revolves around formulas and their meaning, we demonstrate a one of a kind usage of entity cards as a form of auto-complete. This provides an enhanced ecosystem where users can seamlessly lookup and consume factual information, about formulas and mathematical concepts with minimal effort. We propose one approach to address the challenges faced by polysemic symbols, by creating multiple cards based on the concept, and using faceted search to help navigate content more logically and seamlessly.

We conducted a human experiment to observe the usefulness of these cards, in isolation under a math information retrieval setting. This gave us the opportunity to compare the impacts individual components of the cards had on users. The study was designed to accommodate inputs in both text and formula (pre-rendered LaTeX) format, while also controlling for familiarity of a concept and formula size. A key insight from the experiment is for formula only search, providing a description of a concept might not always suffice. Apart from the descriptions of the concept to which the formulas are associated, users are also trying to understand the meaning of the variables and information about the operations that connect them.

7.1 Contributions

Overall the contributions of the thesis are as follows:

1. An alternate design of entity cards specifically meant to address various types of mathematical search needs, that current entity cards for text-based mathematical search do not address.

2. Populating individual components (title, formula and description) of these cards by compiling data from existing structured and semi-structured data-sources.

3. A human experiment to study the usefulness of individual card components while searching for mathematical content from both a text query and a LaTeX query input.

4. Creation of an index on both titles and formulas, that can be queried via an API, and demonstrating an alternate use of these cards as a form of auto-complete.

In the next section we suggest primary areas to focus further research on. We also release our data-set of all formulas, titles, descriptions and aliases to facilitate further research on improvising the extraction process for descriptions and formula linking to titles.

7.2 Future Work

As seen in the results, we notice that math entity cards, although seemingly beneficial can be improved upon. In this section, we discuss future directions of research based on two main criteria improving data quality and additional benefits.

7.2.1 Data Quality

Since no knowledge base is ever complete, we resort to extracting data from multiple sources so as to provide not just more formulas but also more descriptions for existing concepts. This leads us to a new challenge of data consistency, where some descriptions might consists of only a few words and some may have esoteric descriptions.

Since we make use of pre-existing sources along with rules and pattern based matching for extraction of formulas, our formula association might not be 100% accurate. Even so, we notice some incorrect links in Wikidata that makes use of bots and automated process for extraction. Although manual validation would be ideal, it is not practical. We instead propose developing a system that initially classifies the links between formulas and titles or keywords in description against alternate data-sources, this will help filter ones that are incorrect.

In a manner similar to existing text search engines, we could also make use of user feedback on individual card components to figure out those concepts that need correction.

7.2.2 Additional Benefits

7.2.2.1 Use of Computer Algebra System

With the help of a computer algebra system (CAS) such as Maple[1], Mathematica [2], or SageMath [3], it is possible to identify factored forms of formulas, inverses that can then be used to either find or suggest mathematical entity cards that relate to the formula. For example, if while solving a particular equation, the system narrows down to a quadratic equation of the form $ax^2 + bx + c = 0$ the system could then suggest a card for the quadratic equation, this would help a user to learn and recollect concepts during the solving process. Another way a CAS would be beneficial would be for formulas that are not yet handled by the unification of variables and operators for search e.g., x^{-1} and $\frac{1}{x}$ both represent the concept of an inverse, which can be identified by a CAS. This way irrespective of the input a Math Entity Card for the Inverse could be suggested.

7.2.2.2 Tutorial Links & Related Work

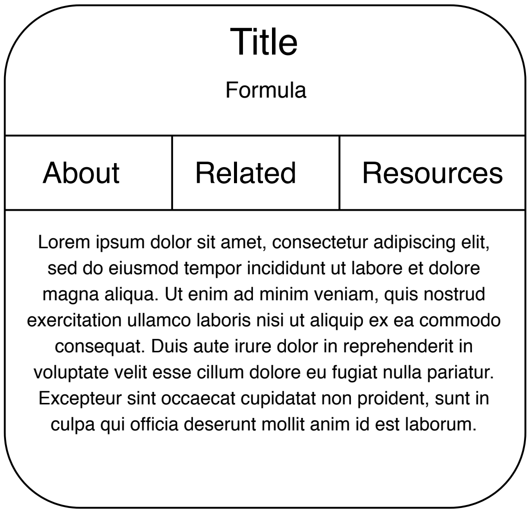
Figure 7.1 demonstrates what we think would be a complete math entity card design. There are three tabs, of ‘About’ - containing the single/multiple descriptions, ‘Related’ - for examples and/or related concepts, ‘Resources’ - for additional resources on understanding the current concept.

Jiang et al.[13] had demonstrated a usage of a PDF Reader with Math-

[1] https://www.maplesoft.com/

[2] https://www.wolfram.com/mathematica/

[3] http://www.sagemath.org/



Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Figure 7.1: A complete math entity card design

Assistant (PRMA) that could recommend Open Educational Resources (OERs), e.g., video, Wikipedia page, or slides to users. A similar approach could be tailored to create suggestions for math entity cards. Along with this approach, we believe capturing the clicks of users could be utilized to tailor the suggestions for resource and related concept on the card. Related concepts could also be mined from existing ‘See-Also’ section on pages from Wikipedia.

As seen earlier, for math information retrieval math entity cards act as an interesting piece of the navigation puzzle between Formulas and Concepts. They help address a factual informational need from both ends: users searching for the names and description of a formula, as well as users searching for the representative formula for a particular concept. Math entity cards help in this bidirectional access of information, without increasing the overall need to filter through more information.

Appendices

Appendix A

Recruiting Email

To: XXXXX

Subject: Seeking Participants for Math Search Experiment

The Document and Pattern Recognition Lab (DPRL) at RIT is seeking participants for an experiment studying new math-aware search engines. These search engines compare documents using both their text and math, and support search using queries that contain keywords and formulas.

The study should last 30 minutes. Participants will be paid $10 for their time.

If you would like to participate in the project or have any questions, please contact Abishai Dmello (ad7527@rit.edu).

Questions about your rights as a participant may be directed to Heather Foti (Associate Director, Human Subjects Research Office, RIT: hmfsrs@rit.edu (585) 475-7673) or Dr. Zanibbi (Principal Investigator, rxzvcs@rit.edu, (585) 475-5023).

Thank you for your time.

Sincerely,

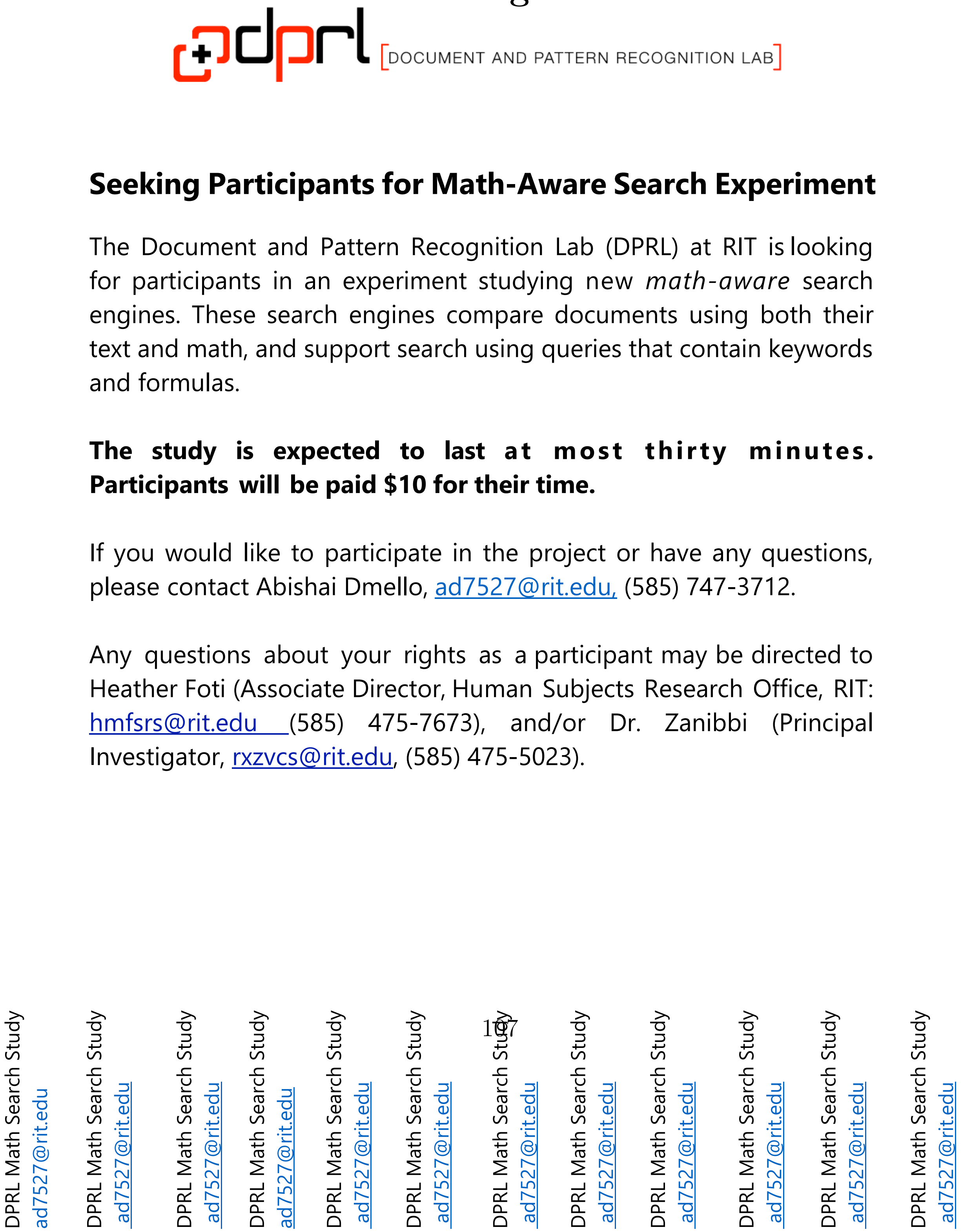
Dr. Richard Zanibbi

Professor, Department of Computer Science,

RIT DPRL Web Page: http://www.cs.rit.edu/ dprl

Appendix B

Recruiting Poster



DOCUMENT AND PATTERN RECOGNITION LAB

Seeking Participants for Math-Aware Search Experiment

The Document and Pattern Recognition Lab (DPRL) at RIT is looking for participants in an experiment studying new *math-aware* search engines. These search engines compare documents using both their text and math, and support search using queries that contain keywords and formulas.

The study is expected to last at most thirty minutes. Participants will be paid $10 for their time.

If you would like to participate in the project or have any questions, please contact Abishai Dmello, ad7527@rit.edu, (585) 747-3712.

Any questions about your rights as a participant may be directed to Heather Foti (Associate Director, Human Subjects Research Office, RIT: hmfsrs@rit.edu (585) 475-7673), and/or Dr. Zanibbi (Principal Investigator, rxzvcs@rit.edu, (585) 475-5023).

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

DPRL Math Search Study
ad7527@rit.edu

Appendix C

Pre-Written Script

Thank you for volunteering to participate in our user study.

The study is being conducted to understand, information preference in users performing Math Information Retrieval.

No personally identifiable information is collected during the study, the system will assign you a randomly generated ID. Your name and University ID is collected for university financial purposes and will not be included in any reports or further publications of the data.

The study is expected to take about 30-35 min to complete, You are free to leave the study at any point in time if you feel uncomfortable. You will be reimbursed at the end of the study or at any point you leave, provided you have signed the consent form.

— Time to Read and Sign Consent Form —

Instructions :

- We are going to show you a series of queries, in the form of text-keywords or math formulas.

- You would then be asked to assess cards, intended to provide information related to the queries.

A reminder :

- The evaluation is purely of the system under test and is in no way intended to serve as a test of your mathematical knowledge.
- You are encouraged to take your time to carefully consider each scenario before responding, but do respond as quickly as possible as it is a timed task.

The experiment has the following sections:

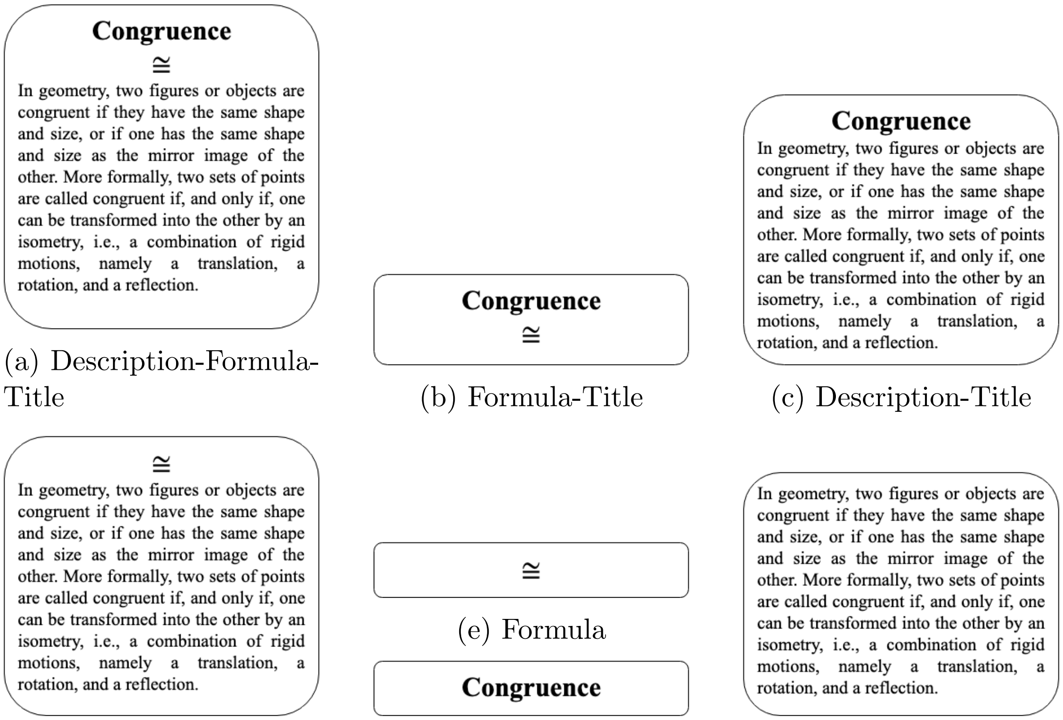
- Instruction Page
- Demographic Survey
- Practice Trials : You would have four practice trials to help familiarize yourself with the interface. Feel free to ask questions, if any during the practice trials, after which no questions can be answered. You are encouraged to respond making your best judgement.
- User Study
- Post-Study Questionnaire.

You are free to change your responses before proceeding to the next section or the next question. Please do not use the browser back button, or any keyboard shortcuts to go to the previous page at any point of the study.

Appendix D

Card Types in Human Experiment

D.1 Symbols



Congruence

$\cong$

In geometry, two figures or objects are congruent if they have the same shape and size, or if one has the same shape and size as the mirror image of the other. More formally, two sets of points are called congruent if, and only if, one can be transformed into the other by an isometry, i.e., a combination of rigid motions, namely a translation, a rotation, and a reflection.

(a) Description-Formula-Title

Congruence

$\cong$

(b) Formula-Title

Congruence

In geometry, two figures or objects are congruent if they have the same shape and size, or if one has the same shape and size as the mirror image of the other. More formally, two sets of points are called congruent if, and only if, one can be transformed into the other by an isometry, i.e., a combination of rigid motions, namely a translation, a rotation, and a reflection.

(c) Description-Title

$\cong$

In geometry, two figures or objects are congruent if they have the same shape and size, or if one has the same shape and size as the mirror image of the other. More formally, two sets of points are called congruent if, and only if, one can be transformed into the other by an isometry, i.e., a combination of rigid motions, namely a translation, a rotation, and a reflection.

(d) Description-Formula

$\cong$

(e) Formula

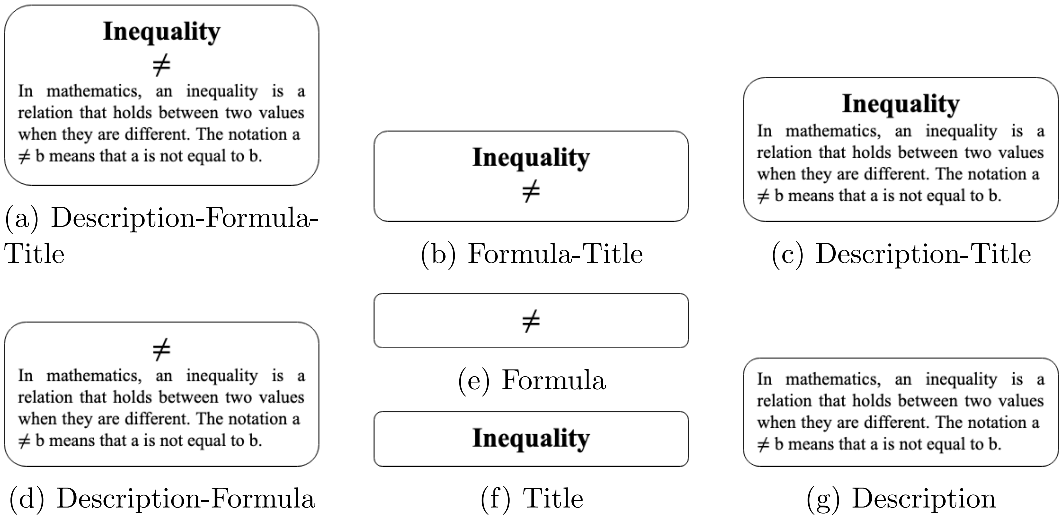
Congruence

(f) Title

In geometry, two figures or objects are congruent if they have the same shape and size, or if one has the same shape and size as the mirror image of the other. More formally, two sets of points are called congruent if, and only if, one can be transformed into the other by an isometry, i.e., a combination of rigid motions, namely a translation, a rotation, and a reflection.

(g) Description

Figure D.1: Card Types for Congruence



Inequality

$\neq$

In mathematics, an inequality is a relation that holds between two values when they are different. The notation a ≠ b means that a is not equal to b.

(a) Description-Formula-Title

Inequality

$\neq$

(b) Formula-Title

Inequality

In mathematics, an inequality is a relation that holds between two values when they are different. The notation a ≠ b means that a is not equal to b.

(c) Description-Title

$\neq$

In mathematics, an inequality is a relation that holds between two values when they are different. The notation a ≠ b means that a is not equal to b.

(d) Description-Formula

$\neq$

(e) Formula

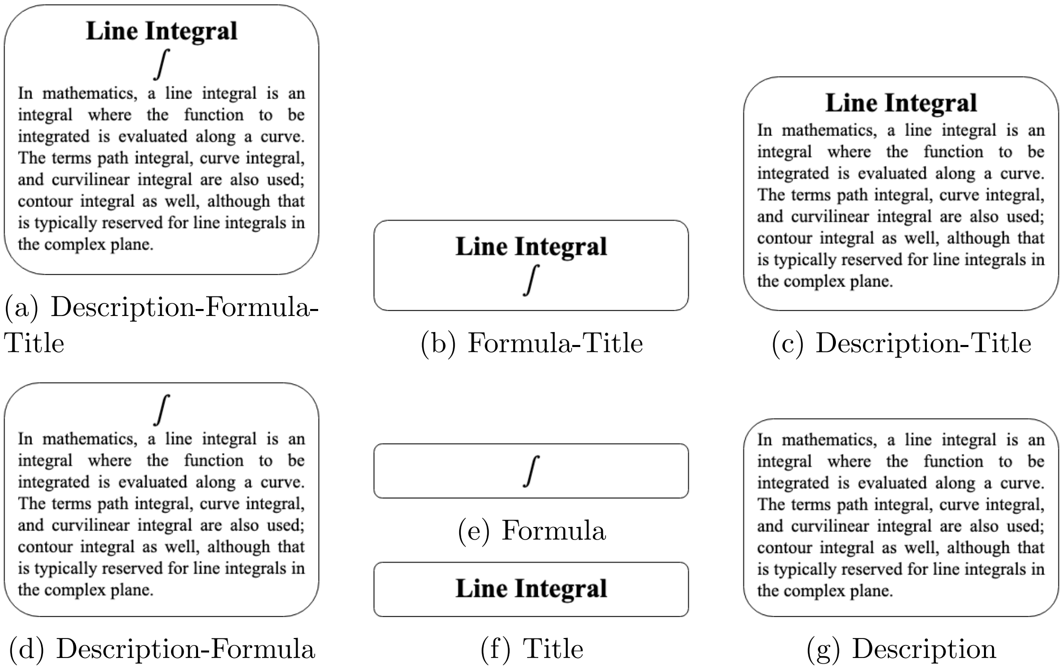
Inequality

(f) Title

In mathematics, an inequality is a relation that holds between two values when they are different. The notation a ≠ b means that a is not equal to b.

(g) Description

Figure D.2: Card Types for Inequality



Line Integral

$\int$

In mathematics, a line integral is an integral where the function to be integrated is evaluated along a curve. The terms path integral, curve integral, and curvilinear integral are also used; contour integral as well, although that is typically reserved for line integrals in the complex plane.

(a) Description-Formula-Title

Line Integral

$\int$

(b) Formula-Title

Line Integral

In mathematics, a line integral is an integral where the function to be integrated is evaluated along a curve. The terms path integral, curve integral, and curvilinear integral are also used; contour integral as well, although that is typically reserved for line integrals in the complex plane.

(c) Description-Title

$\int$

In mathematics, a line integral is an integral where the function to be integrated is evaluated along a curve. The terms path integral, curve integral, and curvilinear integral are also used; contour integral as well, although that is typically reserved for line integrals in the complex plane.

(d) Description-Formula

$\int$

(e) Formula

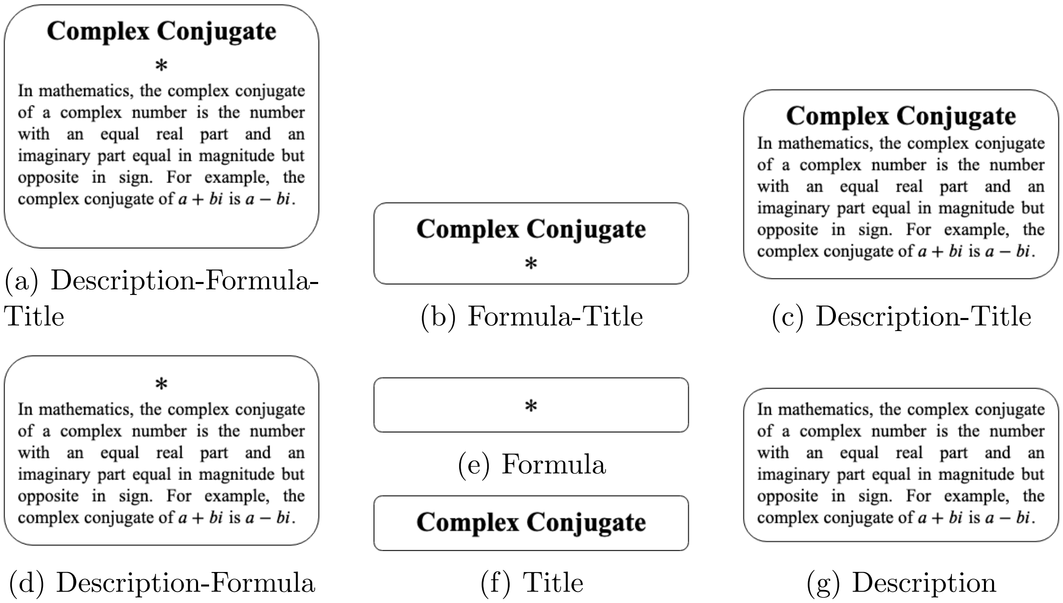
Line Integral

(f) Title

In mathematics, a line integral is an integral where the function to be integrated is evaluated along a curve. The terms path integral, curve integral, and curvilinear integral are also used; contour integral as well, although that is typically reserved for line integrals in the complex plane.

(g) Description

Figure D.3: Card Types for Line Integral


Complex Conjugate
*
In mathematics, the complex conjugate of a complex number is the number with an equal real part and an imaginary part equal in magnitude but opposite in sign. For example, the complex conjugate of $a + bi$ is $a - bi$.

(a) Description-Formula-Title

(b) Formula-Title

(c) Description-Title

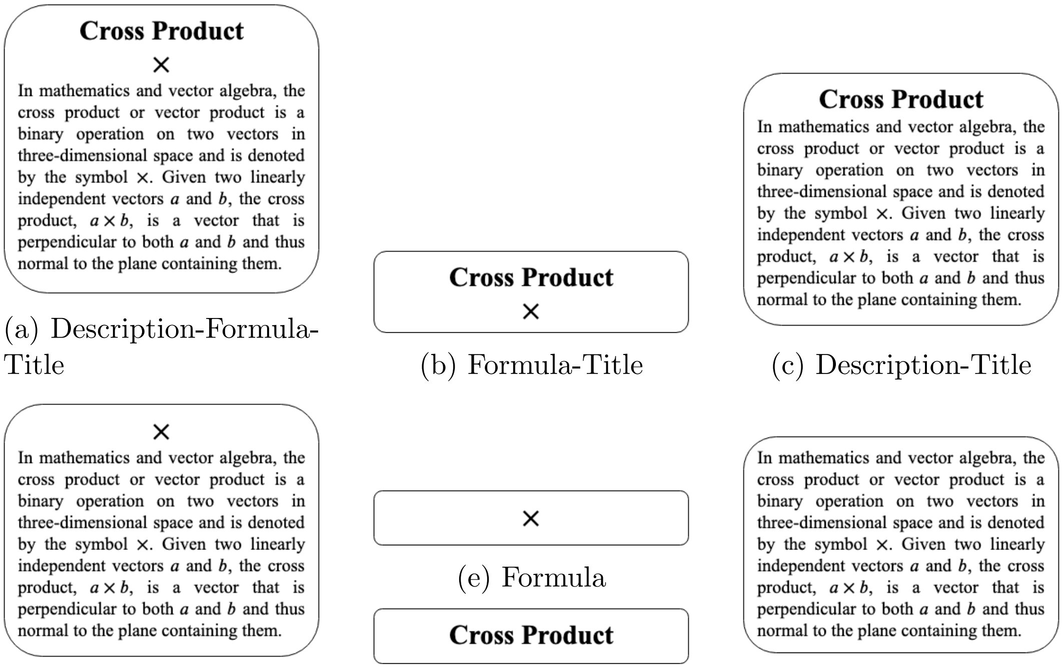
(d) Description-Formula

(e) Formula

(f) Title

(g) Description

Figure D.4: Card Types for Complex Conjugate



Cross Product

×

In mathematics and vector algebra, the cross product or vector product is a binary operation on two vectors in three-dimensional space and is denoted by the symbol ×. Given two linearly independent vectors *a* and *b*, the cross product, $a \times b$, is a vector that is perpendicular to both *a* and *b* and thus normal to the plane containing them.

(a) Description-Formula-Title

Cross Product

×

(b) Formula-Title

Cross Product

In mathematics and vector algebra, the cross product or vector product is a binary operation on two vectors in three-dimensional space and is denoted by the symbol ×. Given two linearly independent vectors *a* and *b*, the cross product, $a \times b$, is a vector that is perpendicular to both *a* and *b* and thus normal to the plane containing them.

(c) Description-Title

×

In mathematics and vector algebra, the cross product or vector product is a binary operation on two vectors in three-dimensional space and is denoted by the symbol ×. Given two linearly independent vectors *a* and *b*, the cross product, $a \times b$, is a vector that is perpendicular to both *a* and *b* and thus normal to the plane containing them.

(d) Description-Formula

×

(e) Formula

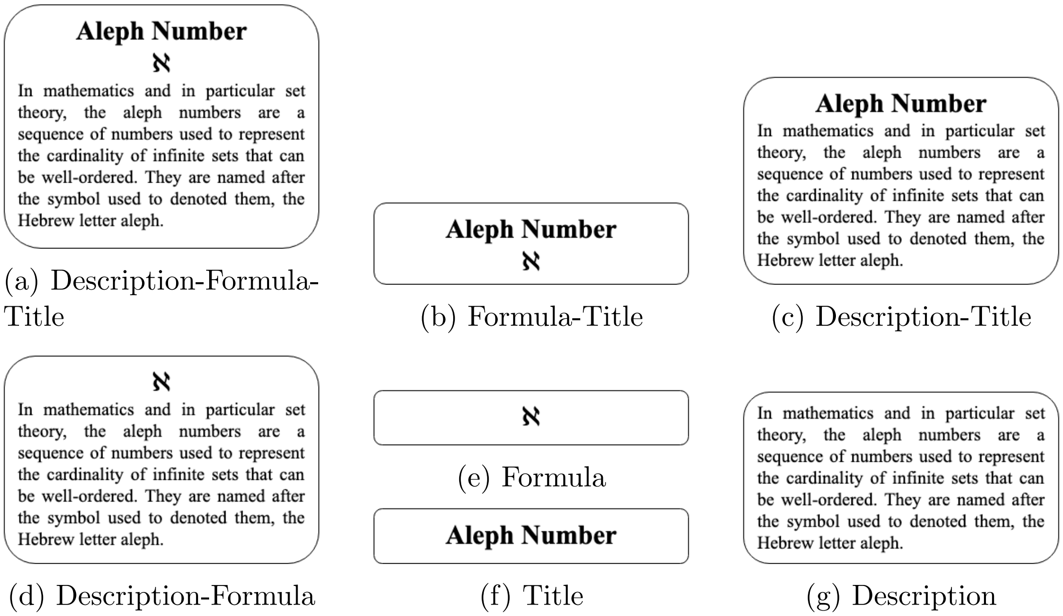
Cross Product

(f) Title

In mathematics and vector algebra, the cross product or vector product is a binary operation on two vectors in three-dimensional space and is denoted by the symbol ×. Given two linearly independent vectors *a* and *b*, the cross product, $a \times b$, is a vector that is perpendicular to both *a* and *b* and thus normal to the plane containing them.

(g) Description

Figure D.5: Card Types for Cross Product



Aleph Number

$\aleph$

In mathematics and in particular set theory, the aleph numbers are a sequence of numbers used to represent the cardinality of infinite sets that can be well-ordered. They are named after the symbol used to denoted them, the Hebrew letter aleph.

(a) Description-Formula-Title

Aleph Number

$\aleph$

(b) Formula-Title

Aleph Number

In mathematics and in particular set theory, the aleph numbers are a sequence of numbers used to represent the cardinality of infinite sets that can be well-ordered. They are named after the symbol used to denoted them, the Hebrew letter aleph.

(c) Description-Title

$\aleph$

In mathematics and in particular set theory, the aleph numbers are a sequence of numbers used to represent the cardinality of infinite sets that can be well-ordered. They are named after the symbol used to denoted them, the Hebrew letter aleph.

(d) Description-Formula

$\aleph$

(e) Formula

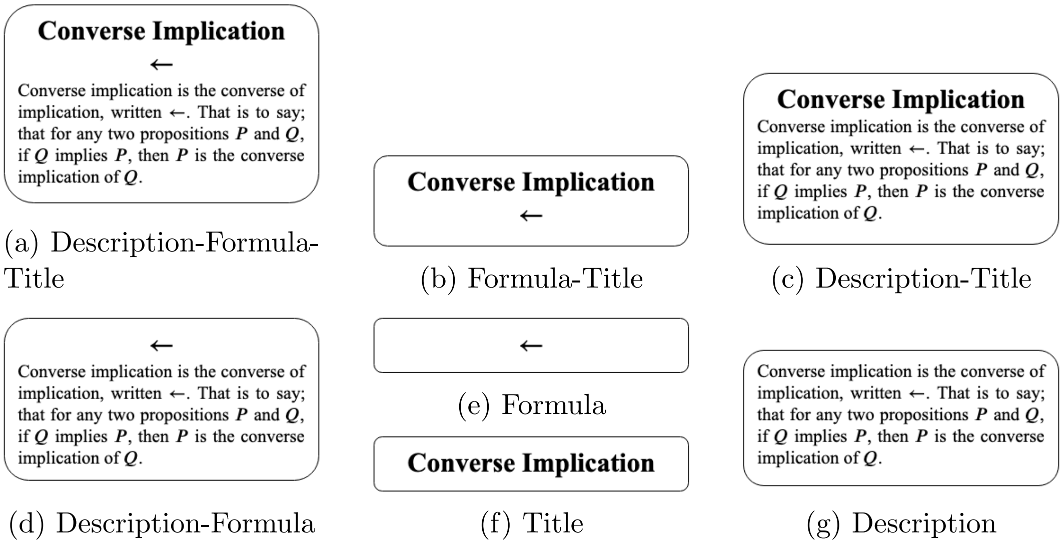
Aleph Number

(f) Title

In mathematics and in particular set theory, the aleph numbers are a sequence of numbers used to represent the cardinality of infinite sets that can be well-ordered. They are named after the symbol used to denoted them, the Hebrew letter aleph.

(g) Description

Figure D.6: Card Types for Aleph Number



Converse Implication

←

Converse implication is the converse of implication, written ←. That is to say; that for any two propositions P and Q, if Q implies P, then P is the converse implication of Q.

(a) Description-Formula-Title

Converse Implication

←

(b) Formula-Title

Converse Implication

Converse implication is the converse of implication, written ←. That is to say; that for any two propositions P and Q, if Q implies P, then P is the converse implication of Q.

(c) Description-Title

←

Converse implication is the converse of implication, written ←. That is to say; that for any two propositions P and Q, if Q implies P, then P is the converse implication of Q.

(d) Description-Formula

←

(e) Formula

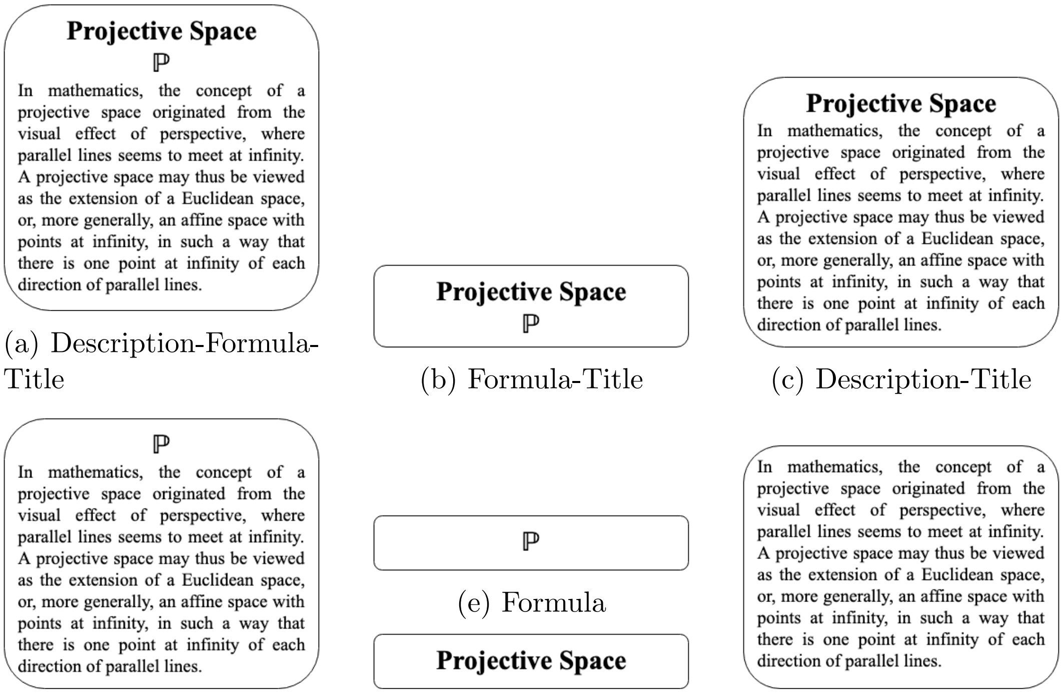
Converse Implication

(f) Title

Converse implication is the converse of implication, written ←. That is to say; that for any two propositions P and Q, if Q implies P, then P is the converse implication of Q.

(g) Description

Figure D.7: Card Types for Converse Implication



Projective Space

$\mathbb{P}$

In mathematics, the concept of a projective space originated from the visual effect of perspective, where parallel lines seems to meet at infinity. A projective space may thus be viewed as the extension of a Euclidean space, or, more generally, an affine space with points at infinity, in such a way that there is one point at infinity of each direction of parallel lines.

(a) Description-Formula-Title

Projective Space

$\mathbb{P}$

(b) Formula-Title

Projective Space

In mathematics, the concept of a projective space originated from the visual effect of perspective, where parallel lines seems to meet at infinity. A projective space may thus be viewed as the extension of a Euclidean space, or, more generally, an affine space with points at infinity, in such a way that there is one point at infinity of each direction of parallel lines.

(c) Description-Title

$\mathbb{P}$

In mathematics, the concept of a projective space originated from the visual effect of perspective, where parallel lines seems to meet at infinity. A projective space may thus be viewed as the extension of a Euclidean space, or, more generally, an affine space with points at infinity, in such a way that there is one point at infinity of each direction of parallel lines.

(d) Description-Formula

$\mathbb{P}$

(e) Formula

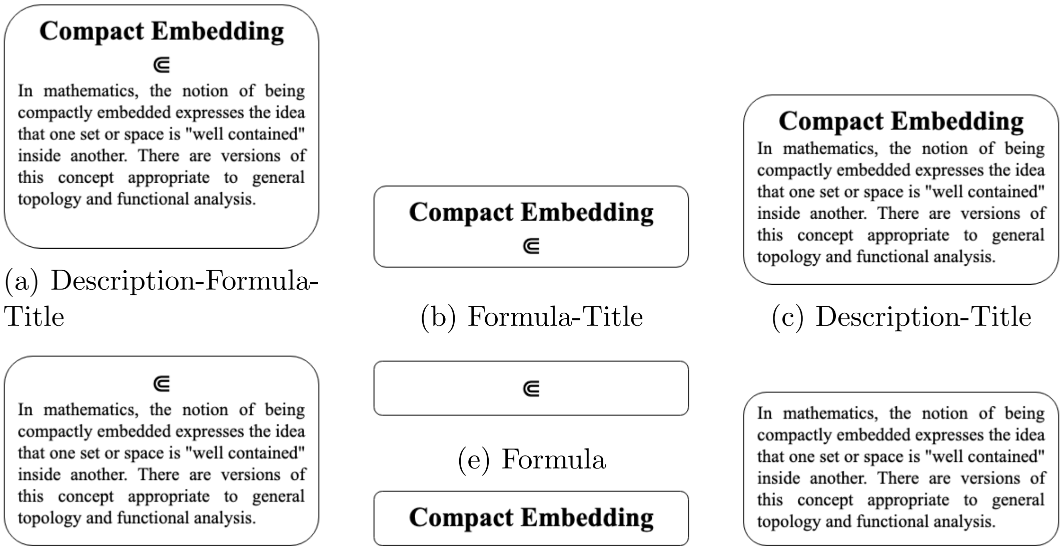
Projective Space

(f) Title

In mathematics, the concept of a projective space originated from the visual effect of perspective, where parallel lines seems to meet at infinity. A projective space may thus be viewed as the extension of a Euclidean space, or, more generally, an affine space with points at infinity, in such a way that there is one point at infinity of each direction of parallel lines.

(g) Description

Figure D.8: Card Types for Projective Space



Compact Embedding

⋐

In mathematics, the notion of being compactly embedded expresses the idea that one set or space is "well contained" inside another. There are versions of this concept appropriate to general topology and functional analysis.

(a) Description-Formula-Title

Compact Embedding

⋐

(b) Formula-Title

Compact Embedding

In mathematics, the notion of being compactly embedded expresses the idea that one set or space is "well contained" inside another. There are versions of this concept appropriate to general topology and functional analysis.

(c) Description-Title

⋐

In mathematics, the notion of being compactly embedded expresses the idea that one set or space is "well contained" inside another. There are versions of this concept appropriate to general topology and functional analysis.

(d) Description-Formula

⋐

(e) Formula

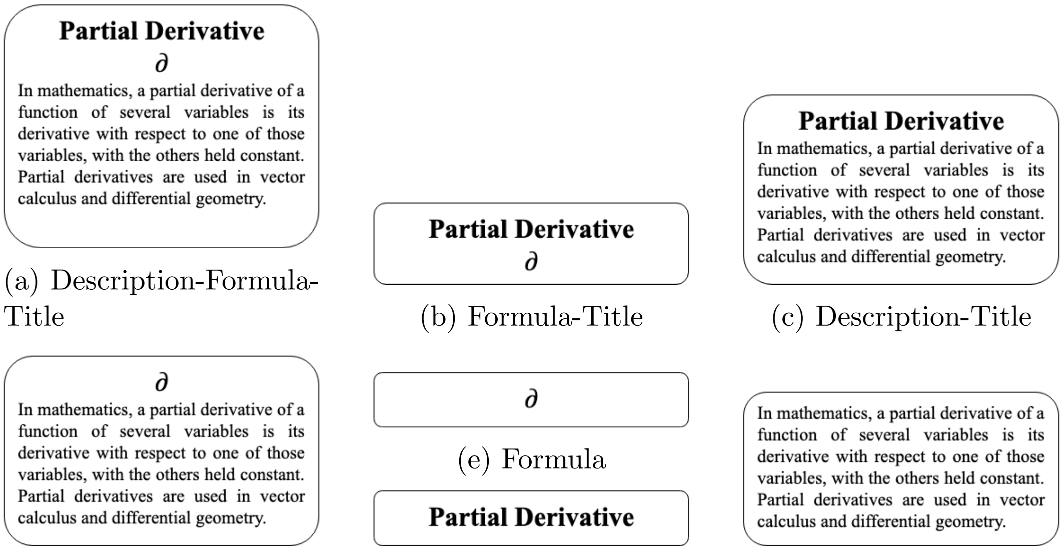
Compact Embedding

(f) Title

In mathematics, the notion of being compactly embedded expresses the idea that one set or space is "well contained" inside another. There are versions of this concept appropriate to general topology and functional analysis.

(g) Description

Figure D.9: Card Types for Compact Embedding



Partial Derivative

∂

In mathematics, a partial derivative of a function of several variables is its derivative with respect to one of those variables, with the others held constant. Partial derivatives are used in vector calculus and differential geometry.

(a) Description-Formula-Title

Partial Derivative

∂

(b) Formula-Title

Partial Derivative

In mathematics, a partial derivative of a function of several variables is its derivative with respect to one of those variables, with the others held constant. Partial derivatives are used in vector calculus and differential geometry.

(c) Description-Title

∂

In mathematics, a partial derivative of a function of several variables is its derivative with respect to one of those variables, with the others held constant. Partial derivatives are used in vector calculus and differential geometry.

(d) Description-Formula

∂

(e) Formula

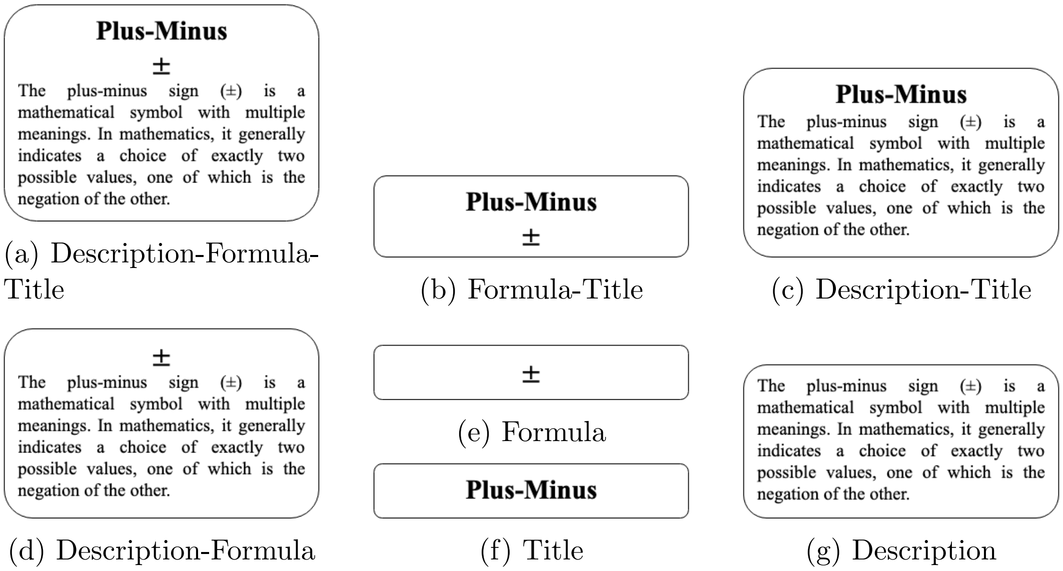
Partial Derivative

(f) Title

In mathematics, a partial derivative of a function of several variables is its derivative with respect to one of those variables, with the others held constant. Partial derivatives are used in vector calculus and differential geometry.

(g) Description

Figure D.10: Card Types for Partial Derivative



Plus-Minus
±
The plus-minus sign (±) is a mathematical symbol with multiple meanings. In mathematics, it generally indicates a choice of exactly two possible values, one of which is the negation of the other.

(a) Description-Formula-Title

Plus-Minus
±

(b) Formula-Title

Plus-Minus
The plus-minus sign (±) is a mathematical symbol with multiple meanings. In mathematics, it generally indicates a choice of exactly two possible values, one of which is the negation of the other.

(c) Description-Title

±
The plus-minus sign (±) is a mathematical symbol with multiple meanings. In mathematics, it generally indicates a choice of exactly two possible values, one of which is the negation of the other.

(d) Description-Formula

±

(e) Formula

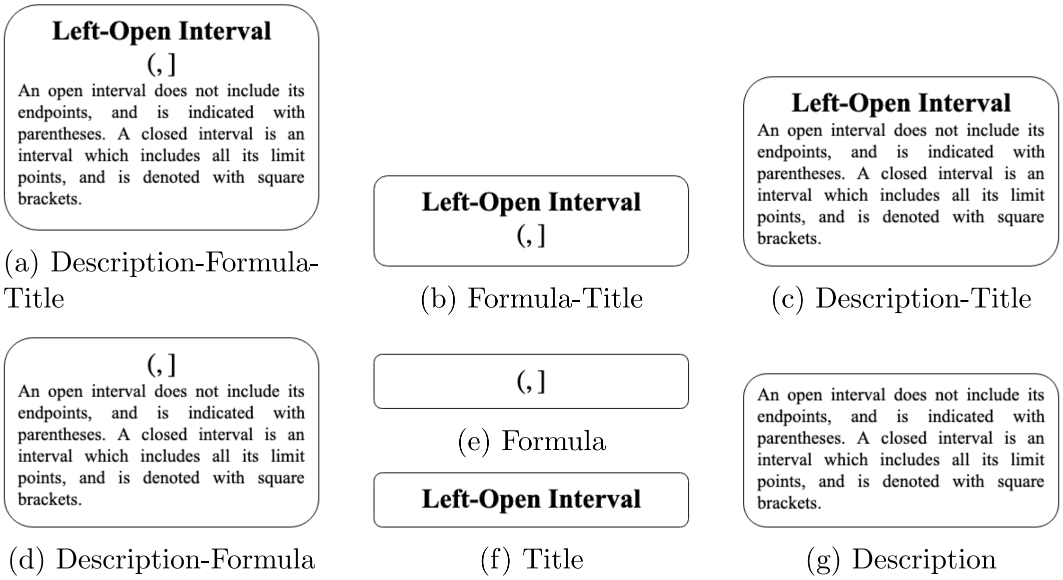
Plus-Minus

(f) Title

The plus-minus sign (±) is a mathematical symbol with multiple meanings. In mathematics, it generally indicates a choice of exactly two possible values, one of which is the negation of the other.

(g) Description

Figure D.11: Card Types for Plus-Minus



Left-Open Interval
(,]
An open interval does not include its endpoints, and is indicated with parentheses. A closed interval is an interval which includes all its limit points, and is denoted with square brackets.

(a) Description-Formula-Title

Left-Open Interval
(,]

(b) Formula-Title

Left-Open Interval
An open interval does not include its endpoints, and is indicated with parentheses. A closed interval is an interval which includes all its limit points, and is denoted with square brackets.

(c) Description-Title

(,]
An open interval does not include its endpoints, and is indicated with parentheses. A closed interval is an interval which includes all its limit points, and is denoted with square brackets.

(d) Description-Formula

(,]

(e) Formula

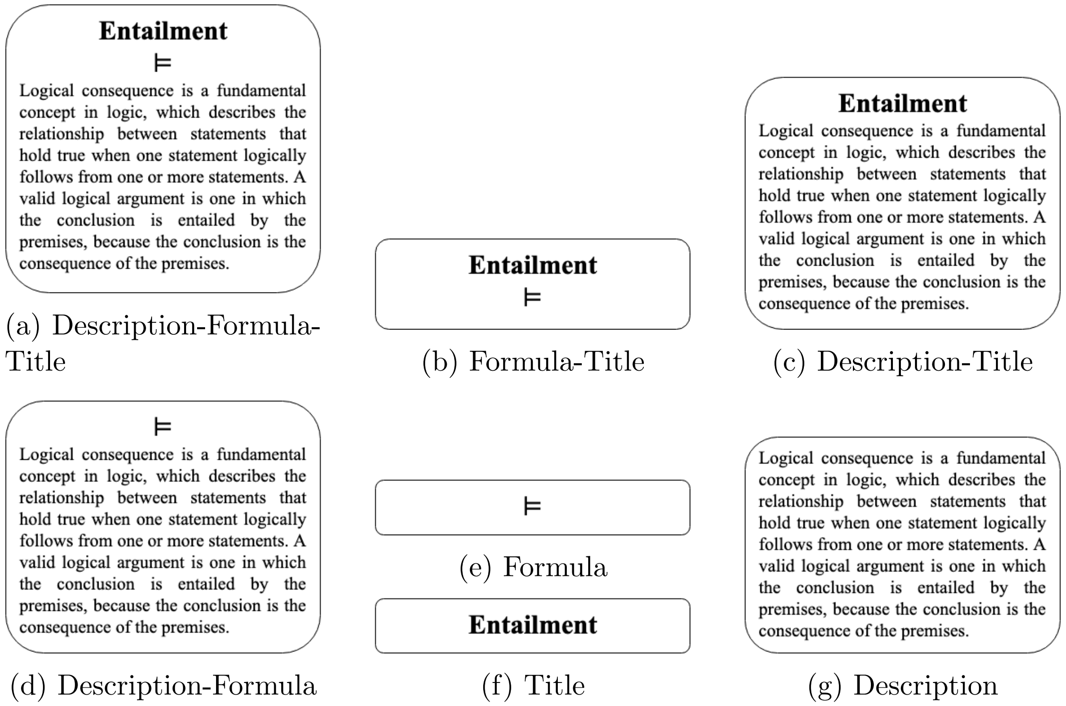
Left-Open Interval

(f) Title

An open interval does not include its endpoints, and is indicated with parentheses. A closed interval is an interval which includes all its limit points, and is denoted with square brackets.

(g) Description

Figure D.12: Card Types for Left Open Interval



Entailment

⊨

Logical consequence is a fundamental concept in logic, which describes the relationship between statements that hold true when one statement logically follows from one or more statements. A valid logical argument is one in which the conclusion is entailed by the premises, because the conclusion is the consequence of the premises.

(a) Description-Formula-Title

Entailment

⊨

(b) Formula-Title

Entailment

Logical consequence is a fundamental concept in logic, which describes the relationship between statements that hold true when one statement logically follows from one or more statements. A valid logical argument is one in which the conclusion is entailed by the premises, because the conclusion is the consequence of the premises.

(c) Description-Title

⊨

Logical consequence is a fundamental concept in logic, which describes the relationship between statements that hold true when one statement logically follows from one or more statements. A valid logical argument is one in which the conclusion is entailed by the premises, because the conclusion is the consequence of the premises.

(d) Description-Formula

⊨

(e) Formula

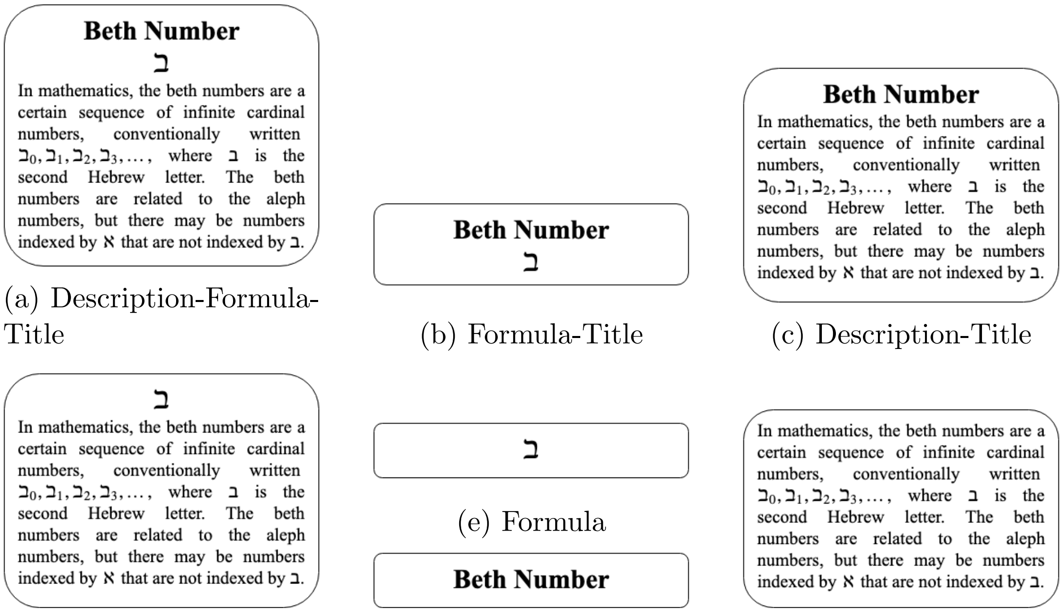
Entailment

(f) Title

Logical consequence is a fundamental concept in logic, which describes the relationship between statements that hold true when one statement logically follows from one or more statements. A valid logical argument is one in which the conclusion is entailed by the premises, because the conclusion is the consequence of the premises.

(g) Description

Figure D.13: Card Types for Entailment



Beth Number

ב

In mathematics, the beth numbers are a certain sequence of infinite cardinal numbers, conventionally written $\beth_0, \beth_1, \beth_2, \beth_3, \ldots$, where ב is the second Hebrew letter. The beth numbers are related to the aleph numbers, but there may be numbers indexed by א that are not indexed by ב.

(a) Description-Formula-Title

Beth Number

ב

(b) Formula-Title

Beth Number

In mathematics, the beth numbers are a certain sequence of infinite cardinal numbers, conventionally written $\beth_0, \beth_1, \beth_2, \beth_3, \ldots$, where ב is the second Hebrew letter. The beth numbers are related to the aleph numbers, but there may be numbers indexed by א that are not indexed by ב.

(c) Description-Title

ב

In mathematics, the beth numbers are a certain sequence of infinite cardinal numbers, conventionally written $\beth_0, \beth_1, \beth_2, \beth_3, \ldots$, where ב is the second Hebrew letter. The beth numbers are related to the aleph numbers, but there may be numbers indexed by א that are not indexed by ב.

(d) Description-Formula

ב

(e) Formula

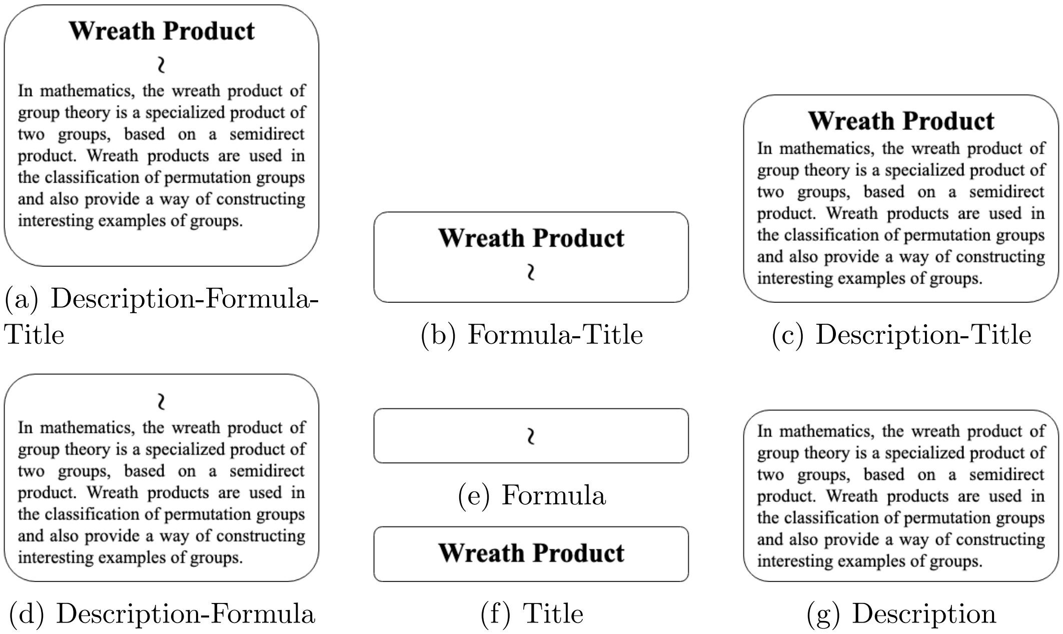
Beth Number

(f) Title

In mathematics, the beth numbers are a certain sequence of infinite cardinal numbers, conventionally written $\beth_0, \beth_1, \beth_2, \beth_3, \ldots$, where ב is the second Hebrew letter. The beth numbers are related to the aleph numbers, but there may be numbers indexed by א that are not indexed by ב.

(g) Description

Figure D.14: Card Types for Beth Number



Wreath Product

≀

In mathematics, the wreath product of group theory is a specialized product of two groups, based on a semidirect product. Wreath products are used in the classification of permutation groups and also provide a way of constructing interesting examples of groups.

(a) Description-Formula-Title

Wreath Product

≀

(b) Formula-Title

Wreath Product

In mathematics, the wreath product of group theory is a specialized product of two groups, based on a semidirect product. Wreath products are used in the classification of permutation groups and also provide a way of constructing interesting examples of groups.

(c) Description-Title

≀

In mathematics, the wreath product of group theory is a specialized product of two groups, based on a semidirect product. Wreath products are used in the classification of permutation groups and also provide a way of constructing interesting examples of groups.

(d) Description-Formula

≀

(e) Formula

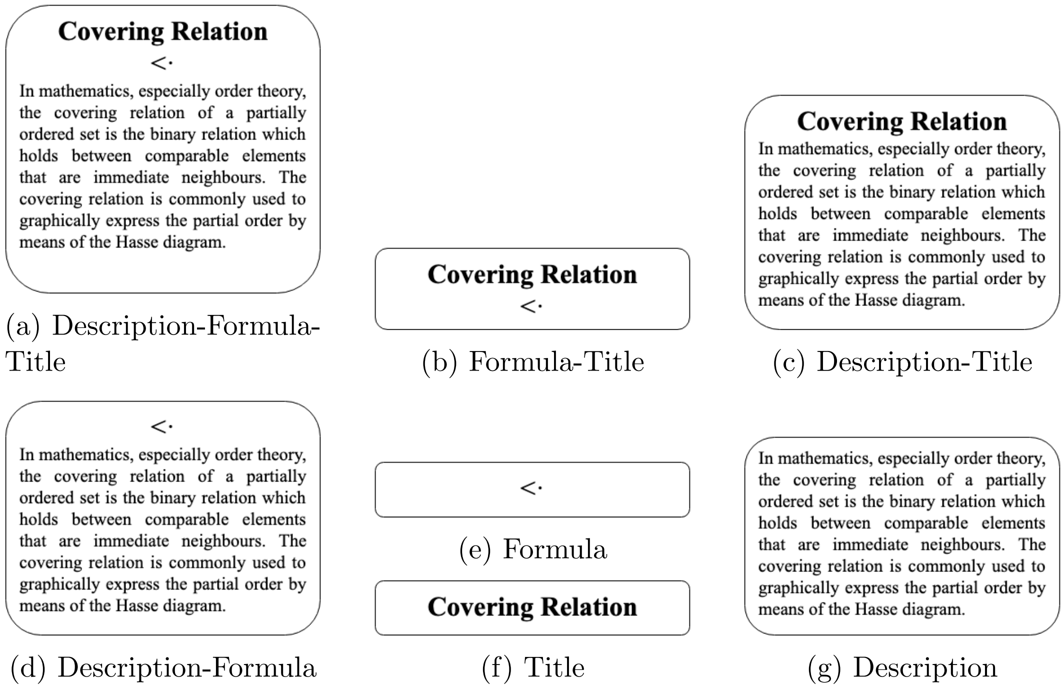
Wreath Product

(f) Title

In mathematics, the wreath product of group theory is a specialized product of two groups, based on a semidirect product. Wreath products are used in the classification of permutation groups and also provide a way of constructing interesting examples of groups.

(g) Description

Figure D.15: Card Types for Wreath Product



Covering Relation

<·

In mathematics, especially order theory, the covering relation of a partially ordered set is the binary relation which holds between comparable elements that are immediate neighbours. The covering relation is commonly used to graphically express the partial order by means of the Hasse diagram.

(a) Description-Formula-Title

Covering Relation

<·

(b) Formula-Title

Covering Relation

In mathematics, especially order theory, the covering relation of a partially ordered set is the binary relation which holds between comparable elements that are immediate neighbours. The covering relation is commonly used to graphically express the partial order by means of the Hasse diagram.

(c) Description-Title

<·

In mathematics, especially order theory, the covering relation of a partially ordered set is the binary relation which holds between comparable elements that are immediate neighbours. The covering relation is commonly used to graphically express the partial order by means of the Hasse diagram.

(d) Description-Formula

<·

(e) Formula

Covering Relation

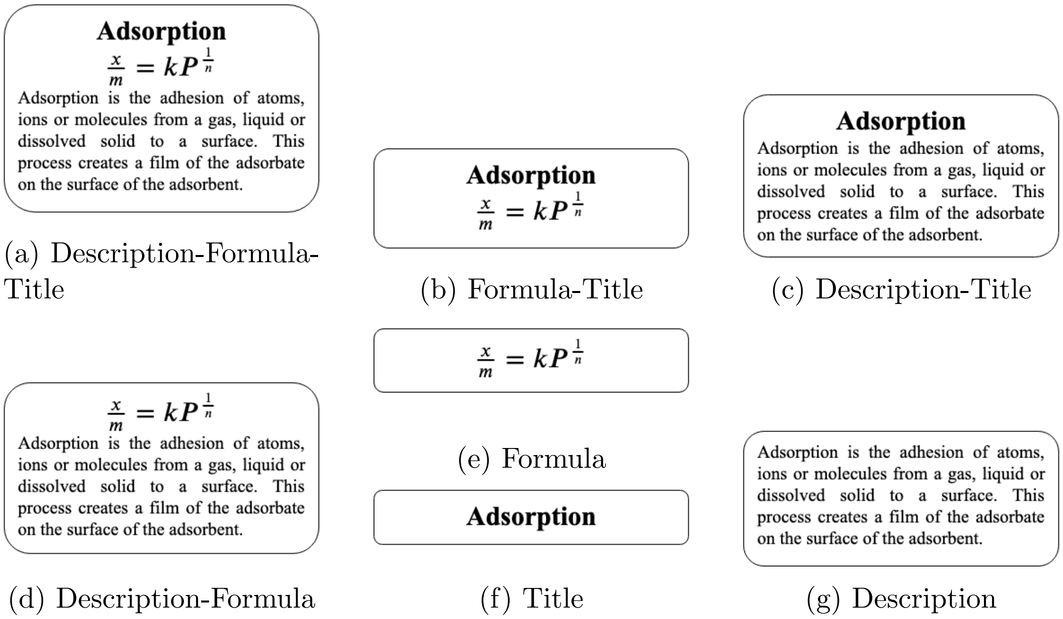
(f) Title

In mathematics, especially order theory, the covering relation of a partially ordered set is the binary relation which holds between comparable elements that are immediate neighbours. The covering relation is commonly used to graphically express the partial order by means of the Hasse diagram.

(g) Description

Figure D.16: Card Types for Covering Relation

D.2 Small Formulas



Adsorption

$\frac{x}{m} = kP^{\frac{1}{n}}$

Adsorption is the adhesion of atoms, ions or molecules from a gas, liquid or dissolved solid to a surface. This process creates a film of the adsorbate on the surface of the adsorbent.

(a) Description-Formula-Title

Adsorption

$\frac{x}{m} = kP^{\frac{1}{n}}$

(b) Formula-Title

Adsorption

Adsorption is the adhesion of atoms, ions or molecules from a gas, liquid or dissolved solid to a surface. This process creates a film of the adsorbate on the surface of the adsorbent.

(c) Description-Title

$\frac{x}{m} = kP^{\frac{1}{n}}$

Adsorption is the adhesion of atoms, ions or molecules from a gas, liquid or dissolved solid to a surface. This process creates a film of the adsorbate on the surface of the adsorbent.

(d) Description-Formula

$\frac{x}{m} = kP^{\frac{1}{n}}$

(e) Formula

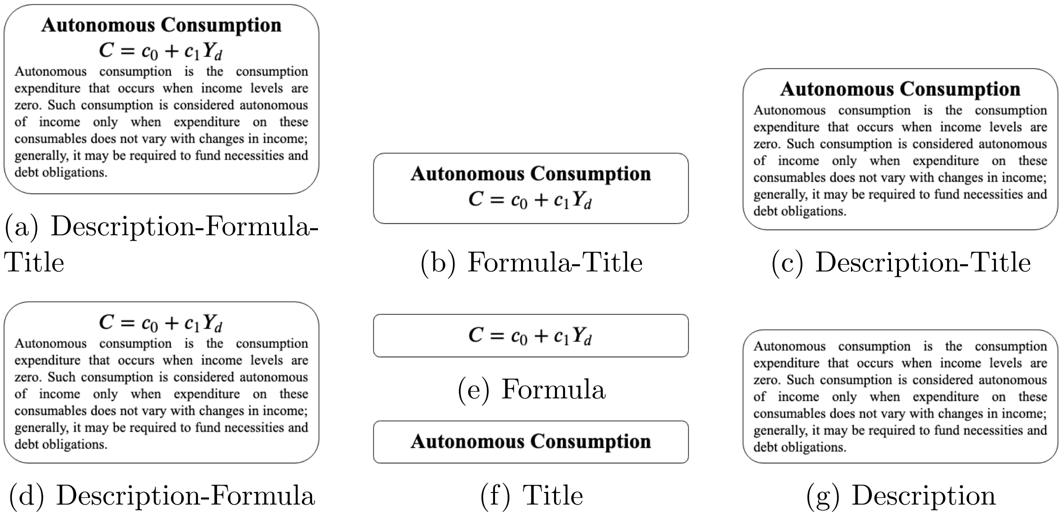
Adsorption

(f) Title

Adsorption is the adhesion of atoms, ions or molecules from a gas, liquid or dissolved solid to a surface. This process creates a film of the adsorbate on the surface of the adsorbent.

(g) Description

Figure D.17: Card Types for Adsorption



Autonomous Consumption

$C = c_0 + c_1 Y_d$

Autonomous consumption is the consumption expenditure that occurs when income levels are zero. Such consumption is considered autonomous of income only when expenditure on these consumables does not vary with changes in income; generally, it may be required to fund necessities and debt obligations.

(a) Description-Formula-Title

Autonomous Consumption

$C = c_0 + c_1 Y_d$

(b) Formula-Title

Autonomous Consumption

Autonomous consumption is the consumption expenditure that occurs when income levels are zero. Such consumption is considered autonomous of income only when expenditure on these consumables does not vary with changes in income; generally, it may be required to fund necessities and debt obligations.

(c) Description-Title

$C = c_0 + c_1 Y_d$

Autonomous consumption is the consumption expenditure that occurs when income levels are zero. Such consumption is considered autonomous of income only when expenditure on these consumables does not vary with changes in income; generally, it may be required to fund necessities and debt obligations.

(d) Description-Formula

$C = c_0 + c_1 Y_d$

(e) Formula

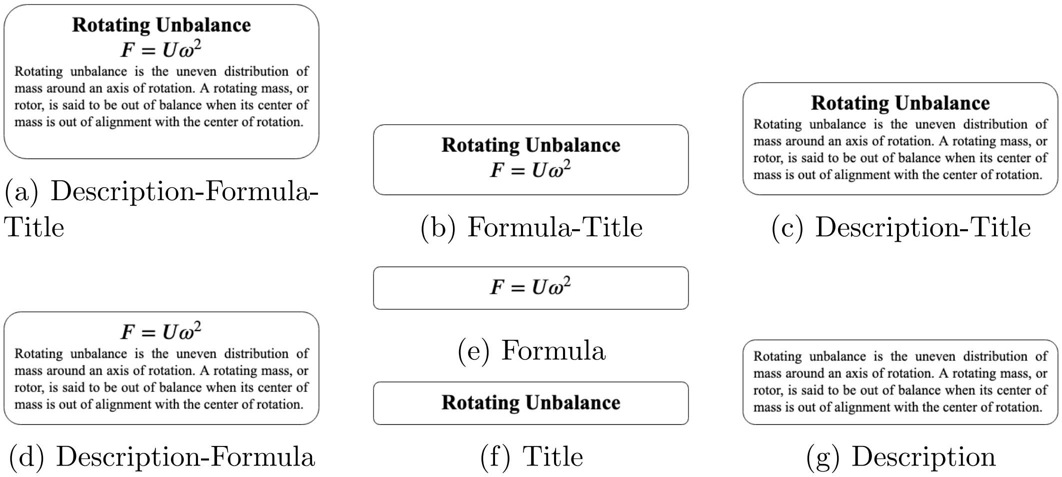
Autonomous Consumption

(f) Title

Autonomous consumption is the consumption expenditure that occurs when income levels are zero. Such consumption is considered autonomous of income only when expenditure on these consumables does not vary with changes in income; generally, it may be required to fund necessities and debt obligations.

(g) Description

Figure D.18: Card Types for Autonomous Consumption



Rotating Unbalance

$F = U\omega^2$

Rotating unbalance is the uneven distribution of mass around an axis of rotation. A rotating mass, or rotor, is said to be out of balance when its center of mass is out of alignment with the center of rotation.

(a) Description-Formula-Title

Rotating Unbalance

$F = U\omega^2$

(b) Formula-Title

Rotating Unbalance

Rotating unbalance is the uneven distribution of mass around an axis of rotation. A rotating mass, or rotor, is said to be out of balance when its center of mass is out of alignment with the center of rotation.

(c) Description-Title

$F = U\omega^2$

Rotating unbalance is the uneven distribution of mass around an axis of rotation. A rotating mass, or rotor, is said to be out of balance when its center of mass is out of alignment with the center of rotation.

(d) Description-Formula

$F = U\omega^2$

(e) Formula

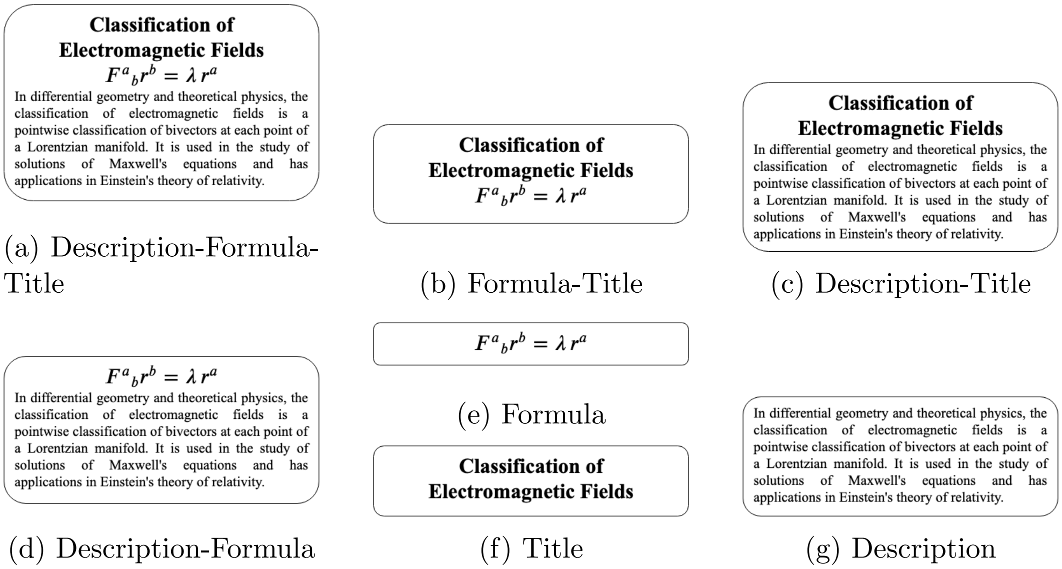
Rotating Unbalance

(f) Title

Rotating unbalance is the uneven distribution of mass around an axis of rotation. A rotating mass, or rotor, is said to be out of balance when its center of mass is out of alignment with the center of rotation.

(g) Description

Figure D.19: Card Types for Rotating Unbalance



Classification of Electromagnetic Fields

$F^a{}_b r^b = \lambda\, r^a$

In differential geometry and theoretical physics, the classification of electromagnetic fields is a pointwise classification of bivectors at each point of a Lorentzian manifold. It is used in the study of solutions of Maxwell's equations and has applications in Einstein's theory of relativity.

(a) Description-Formula-Title

Classification of Electromagnetic Fields

$F^a{}_b r^b = \lambda\, r^a$

(b) Formula-Title

Classification of Electromagnetic Fields

In differential geometry and theoretical physics, the classification of electromagnetic fields is a pointwise classification of bivectors at each point of a Lorentzian manifold. It is used in the study of solutions of Maxwell's equations and has applications in Einstein's theory of relativity.

(c) Description-Title

$F^a{}_b r^b = \lambda\, r^a$

In differential geometry and theoretical physics, the classification of electromagnetic fields is a pointwise classification of bivectors at each point of a Lorentzian manifold. It is used in the study of solutions of Maxwell's equations and has applications in Einstein's theory of relativity.

(d) Description-Formula

$F^a{}_b r^b = \lambda\, r^a$

(e) Formula

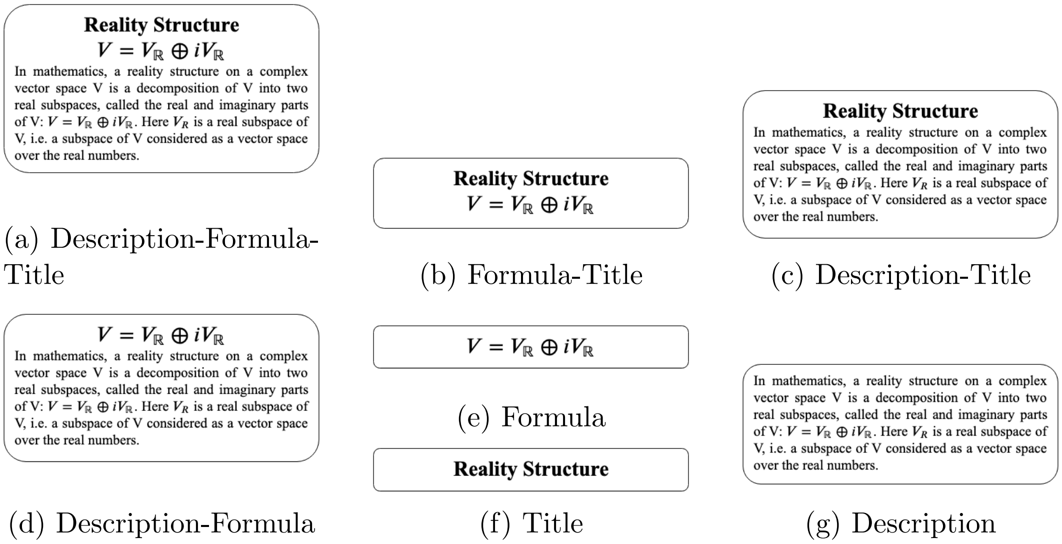
Classification of Electromagnetic Fields

(f) Title

In differential geometry and theoretical physics, the classification of electromagnetic fields is a pointwise classification of bivectors at each point of a Lorentzian manifold. It is used in the study of solutions of Maxwell's equations and has applications in Einstein's theory of relativity.

(g) Description

Figure D.20: Card Types for Classification Of Electromagnetic Fields



Reality Structure

$V = V_{\mathbb{R}} \oplus iV_{\mathbb{R}}$

In mathematics, a reality structure on a complex vector space V is a decomposition of V into two real subspaces, called the real and imaginary parts of V: $V = V_{\mathbb{R}} \oplus iV_{\mathbb{R}}$. Here V_R is a real subspace of V, i.e. a subspace of V considered as a vector space over the real numbers.

(a) Description-Formula-Title

Reality Structure

$V = V_{\mathbb{R}} \oplus iV_{\mathbb{R}}$

(b) Formula-Title

Reality Structure

In mathematics, a reality structure on a complex vector space V is a decomposition of V into two real subspaces, called the real and imaginary parts of V: $V = V_{\mathbb{R}} \oplus iV_{\mathbb{R}}$. Here V_R is a real subspace of V, i.e. a subspace of V considered as a vector space over the real numbers.

(c) Description-Title

$V = V_{\mathbb{R}} \oplus iV_{\mathbb{R}}$

In mathematics, a reality structure on a complex vector space V is a decomposition of V into two real subspaces, called the real and imaginary parts of V: $V = V_{\mathbb{R}} \oplus iV_{\mathbb{R}}$. Here V_R is a real subspace of V, i.e. a subspace of V considered as a vector space over the real numbers.

(d) Description-Formula

$V = V_{\mathbb{R}} \oplus iV_{\mathbb{R}}$

(e) Formula

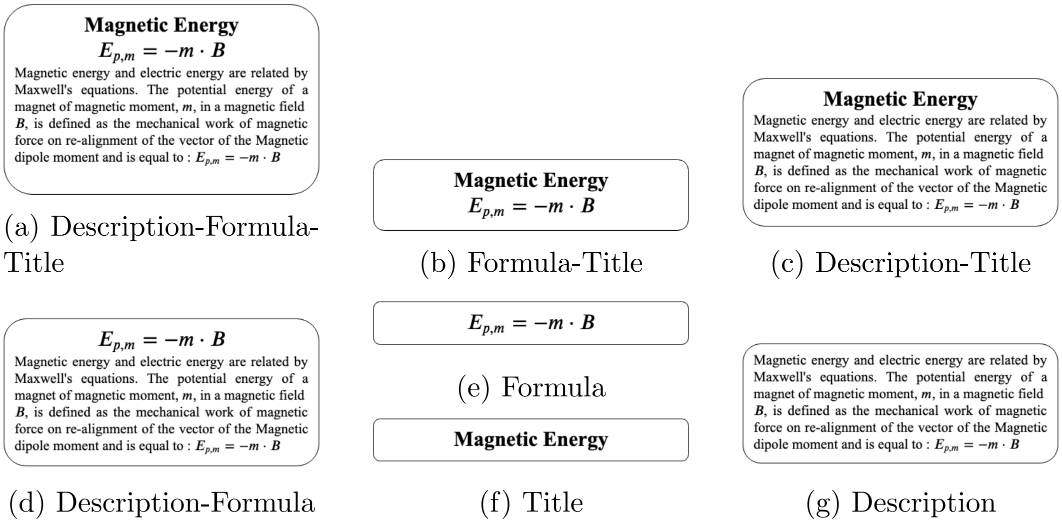
Reality Structure

(f) Title

In mathematics, a reality structure on a complex vector space V is a decomposition of V into two real subspaces, called the real and imaginary parts of V: $V = V_{\mathbb{R}} \oplus iV_{\mathbb{R}}$. Here V_R is a real subspace of V, i.e. a subspace of V considered as a vector space over the real numbers.

(g) Description

Figure D.21: Card Types for Reality Structure



Magnetic Energy

$E_{p,m} = -m \cdot B$

Magnetic energy and electric energy are related by Maxwell's equations. The potential energy of a magnet of magnetic moment, m, in a magnetic field B, is defined as the mechanical work of magnetic force on re-alignment of the vector of the Magnetic dipole moment and is equal to : $E_{p,m} = -m \cdot B$

(a) Description-Formula-Title

Magnetic Energy

$E_{p,m} = -m \cdot B$

(b) Formula-Title

Magnetic Energy

Magnetic energy and electric energy are related by Maxwell's equations. The potential energy of a magnet of magnetic moment, m, in a magnetic field B, is defined as the mechanical work of magnetic force on re-alignment of the vector of the Magnetic dipole moment and is equal to : $E_{p,m} = -m \cdot B$

(c) Description-Title

$E_{p,m} = -m \cdot B$

Magnetic energy and electric energy are related by Maxwell's equations. The potential energy of a magnet of magnetic moment, m, in a magnetic field B, is defined as the mechanical work of magnetic force on re-alignment of the vector of the Magnetic dipole moment and is equal to : $E_{p,m} = -m \cdot B$

(d) Description-Formula

$E_{p,m} = -m \cdot B$

(e) Formula

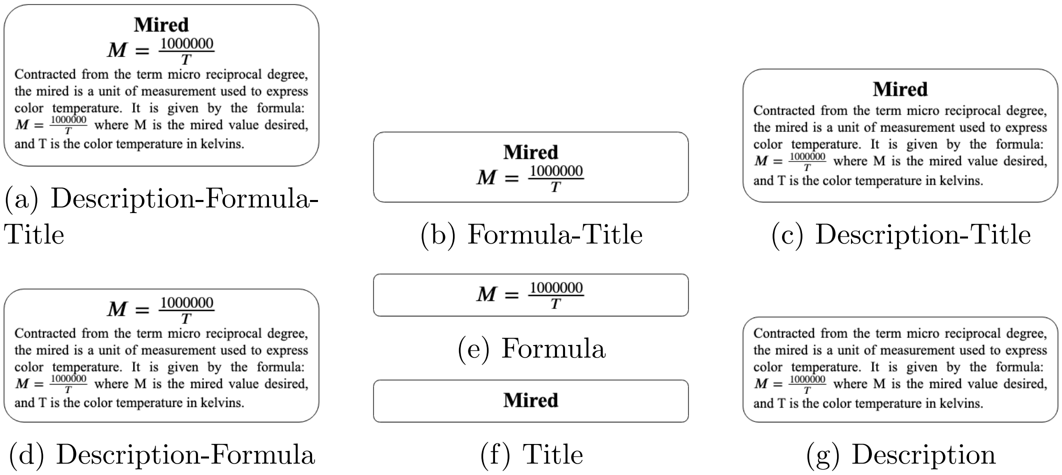
Magnetic Energy

(f) Title

Magnetic energy and electric energy are related by Maxwell's equations. The potential energy of a magnet of magnetic moment, m, in a magnetic field B, is defined as the mechanical work of magnetic force on re-alignment of the vector of the Magnetic dipole moment and is equal to : $E_{p,m} = -m \cdot B$

(g) Description

Figure D.22: Card Types for Magnetic Energy



Mired

$M = \frac{1000000}{T}$

Contracted from the term micro reciprocal degree, the mired is a unit of measurement used to express color temperature. It is given by the formula: $M = \frac{1000000}{T}$ where M is the mired value desired, and T is the color temperature in kelvins.

(a) Description-Formula-Title

Mired

$M = \frac{1000000}{T}$

(b) Formula-Title

Mired

Contracted from the term micro reciprocal degree, the mired is a unit of measurement used to express color temperature. It is given by the formula: $M = \frac{1000000}{T}$ where M is the mired value desired, and T is the color temperature in kelvins.

(c) Description-Title

$M = \frac{1000000}{T}$

Contracted from the term micro reciprocal degree, the mired is a unit of measurement used to express color temperature. It is given by the formula: $M = \frac{1000000}{T}$ where M is the mired value desired, and T is the color temperature in kelvins.

(d) Description-Formula

$M = \frac{1000000}{T}$

(e) Formula

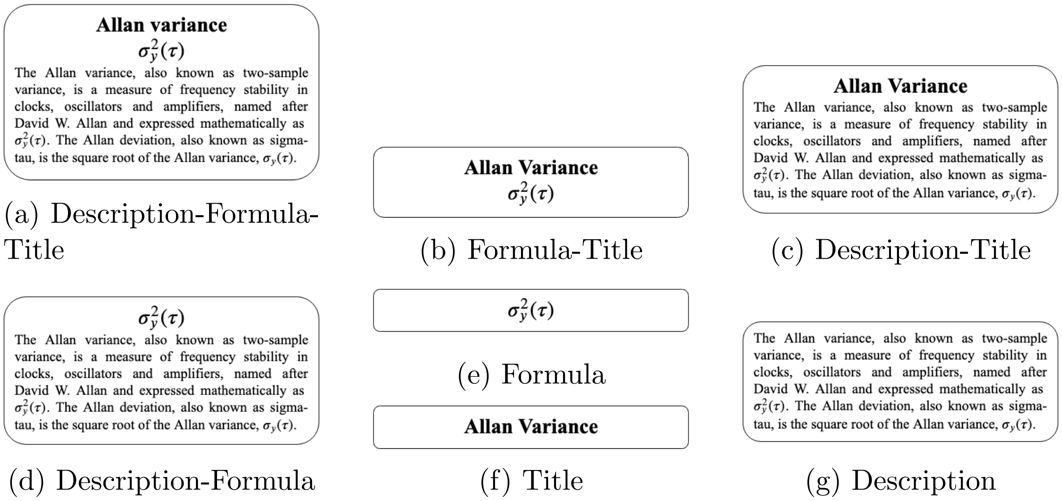
Mired

(f) Title

Contracted from the term micro reciprocal degree, the mired is a unit of measurement used to express color temperature. It is given by the formula: $M = \frac{1000000}{T}$ where M is the mired value desired, and T is the color temperature in kelvins.

(g) Description

Figure D.23: Card Types for Mired



Allan variance

$\sigma_y^2(\tau)$

The Allan variance, also known as two-sample variance, is a measure of frequency stability in clocks, oscillators and amplifiers, named after David W. Allan and expressed mathematically as $\sigma_y^2(\tau)$. The Allan deviation, also known as sigma-tau, is the square root of the Allan variance, $\sigma_y(\tau)$.

(a) Description-Formula-Title

Allan Variance

$\sigma_y^2(\tau)$

(b) Formula-Title

Allan Variance

The Allan variance, also known as two-sample variance, is a measure of frequency stability in clocks, oscillators and amplifiers, named after David W. Allan and expressed mathematically as $\sigma_y^2(\tau)$. The Allan deviation, also known as sigma-tau, is the square root of the Allan variance, $\sigma_y(\tau)$.

(c) Description-Title

$\sigma_y^2(\tau)$

The Allan variance, also known as two-sample variance, is a measure of frequency stability in clocks, oscillators and amplifiers, named after David W. Allan and expressed mathematically as $\sigma_y^2(\tau)$. The Allan deviation, also known as sigma-tau, is the square root of the Allan variance, $\sigma_y(\tau)$.

(d) Description-Formula

$\sigma_y^2(\tau)$

(e) Formula

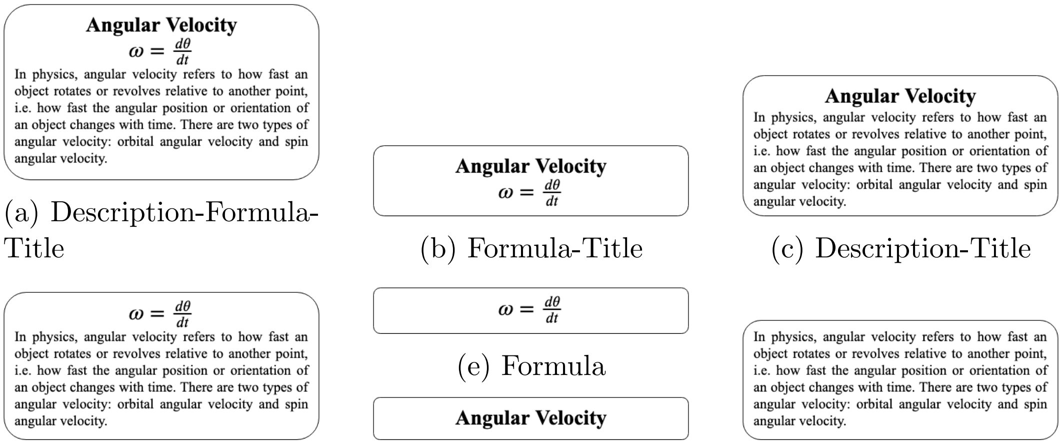
Allan Variance

(f) Title

The Allan variance, also known as two-sample variance, is a measure of frequency stability in clocks, oscillators and amplifiers, named after David W. Allan and expressed mathematically as $\sigma_y^2(\tau)$. The Allan deviation, also known as sigma-tau, is the square root of the Allan variance, $\sigma_y(\tau)$.

(g) Description

Figure D.24: Card Types for Allan Variance



Angular Velocity

$\omega = \frac{d\theta}{dt}$

In physics, angular velocity refers to how fast an object rotates or revolves relative to another point, i.e. how fast the angular position or orientation of an object changes with time. There are two types of angular velocity: orbital angular velocity and spin angular velocity.

(a) Description-Formula-Title

Angular Velocity

$\omega = \frac{d\theta}{dt}$

(b) Formula-Title

Angular Velocity

In physics, angular velocity refers to how fast an object rotates or revolves relative to another point, i.e. how fast the angular position or orientation of an object changes with time. There are two types of angular velocity: orbital angular velocity and spin angular velocity.

(c) Description-Title

$\omega = \frac{d\theta}{dt}$

In physics, angular velocity refers to how fast an object rotates or revolves relative to another point, i.e. how fast the angular position or orientation of an object changes with time. There are two types of angular velocity: orbital angular velocity and spin angular velocity.

(d) Description-Formula

$\omega = \frac{d\theta}{dt}$

(e) Formula

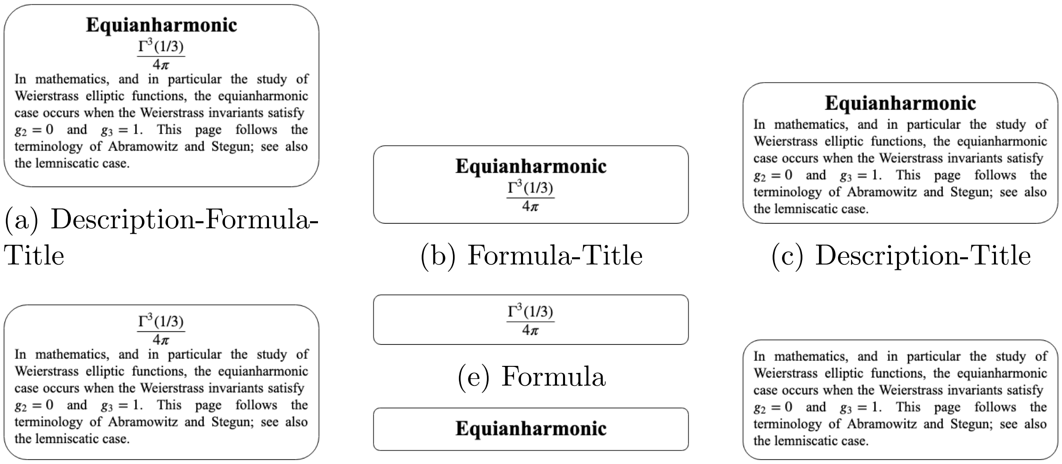
Angular Velocity

(f) Title

In physics, angular velocity refers to how fast an object rotates or revolves relative to another point, i.e. how fast the angular position or orientation of an object changes with time. There are two types of angular velocity: orbital angular velocity and spin angular velocity.

(g) Description

Figure D.25: Card Types for Angular Velocity



Equianharmonic

$\frac{\Gamma^3(1/3)}{4\pi}$

In mathematics, and in particular the study of Weierstrass elliptic functions, the equianharmonic case occurs when the Weierstrass invariants satisfy $g_2 = 0$ and $g_3 = 1$. This page follows the terminology of Abramowitz and Stegun; see also the lemniscatic case.

(a) Description-Formula-Title

Equianharmonic

$\frac{\Gamma^3(1/3)}{4\pi}$

(b) Formula-Title

Equianharmonic

In mathematics, and in particular the study of Weierstrass elliptic functions, the equianharmonic case occurs when the Weierstrass invariants satisfy $g_2 = 0$ and $g_3 = 1$. This page follows the terminology of Abramowitz and Stegun; see also the lemniscatic case.

(c) Description-Title

$\frac{\Gamma^3(1/3)}{4\pi}$

In mathematics, and in particular the study of Weierstrass elliptic functions, the equianharmonic case occurs when the Weierstrass invariants satisfy $g_2 = 0$ and $g_3 = 1$. This page follows the terminology of Abramowitz and Stegun; see also the lemniscatic case.

(d) Description-Formula

$\frac{\Gamma^3(1/3)}{4\pi}$

(e) Formula

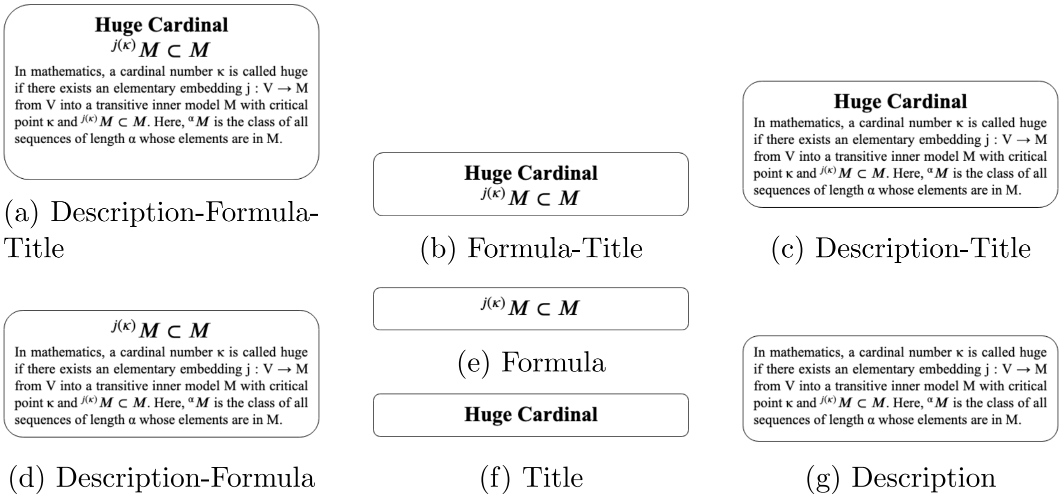
Equianharmonic

(f) Title

In mathematics, and in particular the study of Weierstrass elliptic functions, the equianharmonic case occurs when the Weierstrass invariants satisfy $g_2 = 0$ and $g_3 = 1$. This page follows the terminology of Abramowitz and Stegun; see also the lemniscatic case.

(g) Description

Figure D.26: Card Types for Equianharmonic



Huge Cardinal

$^{j(\kappa)}M \subset M$

In mathematics, a cardinal number κ is called huge if there exists an elementary embedding j : V → M from V into a transitive inner model M with critical point κ and $^{j(\kappa)}M \subset M$. Here, $^{\alpha}M$ is the class of all sequences of length α whose elements are in M.

(a) Description-Formula-Title

Huge Cardinal

$^{j(\kappa)}M \subset M$

(b) Formula-Title

Huge Cardinal

In mathematics, a cardinal number κ is called huge if there exists an elementary embedding j : V → M from V into a transitive inner model M with critical point κ and $^{j(\kappa)}M \subset M$. Here, $^{\alpha}M$ is the class of all sequences of length α whose elements are in M.

(c) Description-Title

$^{j(\kappa)}M \subset M$

In mathematics, a cardinal number κ is called huge if there exists an elementary embedding j : V → M from V into a transitive inner model M with critical point κ and $^{j(\kappa)}M \subset M$. Here, $^{\alpha}M$ is the class of all sequences of length α whose elements are in M.

(d) Description-Formula

$^{j(\kappa)}M \subset M$

(e) Formula

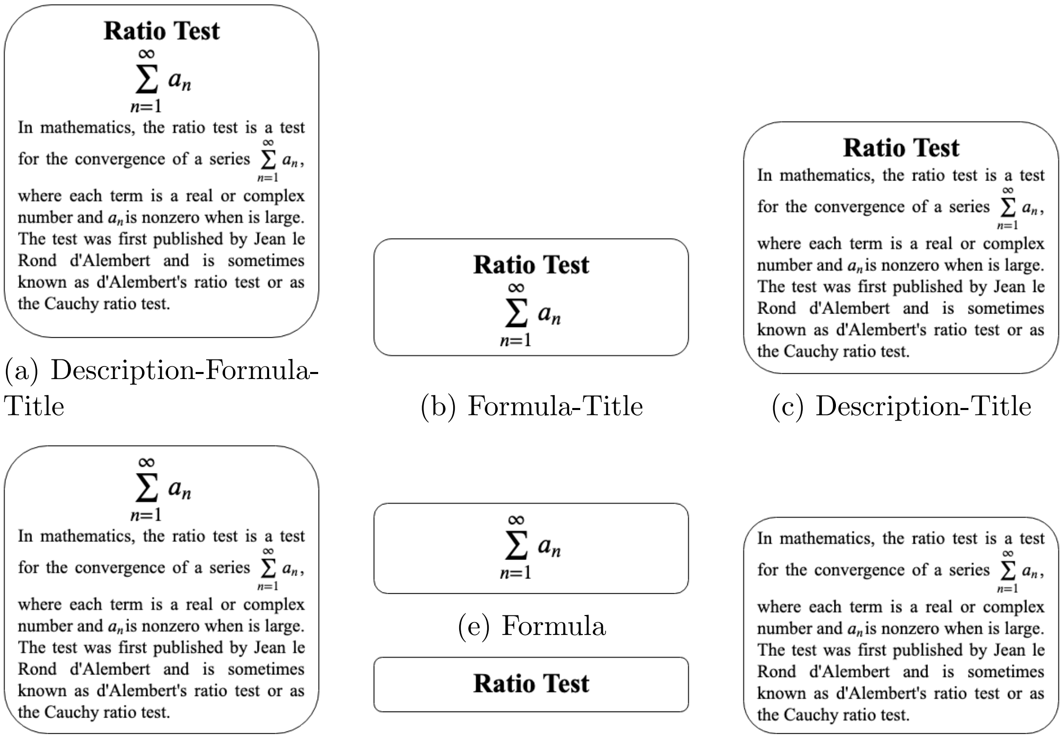
Huge Cardinal

(f) Title

In mathematics, a cardinal number κ is called huge if there exists an elementary embedding j : V → M from V into a transitive inner model M with critical point κ and $^{j(\kappa)}M \subset M$. Here, $^{\alpha}M$ is the class of all sequences of length α whose elements are in M.

(g) Description

Figure D.27: Card Types for Huge Cardinal



Ratio Test

$$\sum_{n=1}^{\infty} a_n$$

In mathematics, the ratio test is a test for the convergence of a series $\sum_{n=1}^{\infty} a_n$, where each term is a real or complex number and a_n is nonzero when is large. The test was first published by Jean le Rond d'Alembert and is sometimes known as d'Alembert's ratio test or as the Cauchy ratio test.

(a) Description-Formula-Title

Ratio Test

$$\sum_{n=1}^{\infty} a_n$$

(b) Formula-Title

Ratio Test

In mathematics, the ratio test is a test for the convergence of a series $\sum_{n=1}^{\infty} a_n$, where each term is a real or complex number and a_n is nonzero when is large. The test was first published by Jean le Rond d'Alembert and is sometimes known as d'Alembert's ratio test or as the Cauchy ratio test.

(c) Description-Title

$$\sum_{n=1}^{\infty} a_n$$

In mathematics, the ratio test is a test for the convergence of a series $\sum_{n=1}^{\infty} a_n$, where each term is a real or complex number and a_n is nonzero when is large. The test was first published by Jean le Rond d'Alembert and is sometimes known as d'Alembert's ratio test or as the Cauchy ratio test.

(d) Description-Formula

$$\sum_{n=1}^{\infty} a_n$$

(e) Formula

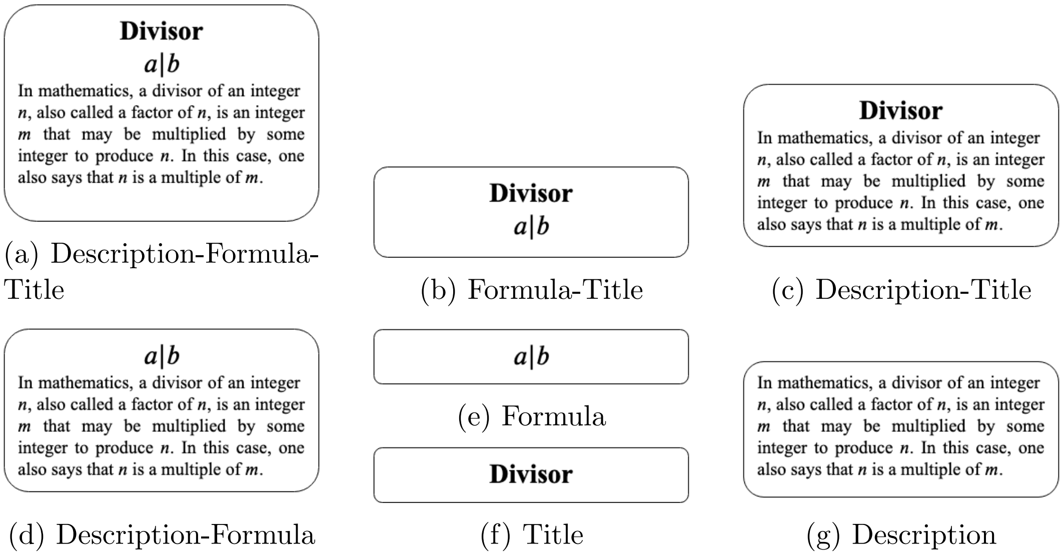
Ratio Test

(f) Title

In mathematics, the ratio test is a test for the convergence of a series $\sum_{n=1}^{\infty} a_n$, where each term is a real or complex number and a_n is nonzero when is large. The test was first published by Jean le Rond d'Alembert and is sometimes known as d'Alembert's ratio test or as the Cauchy ratio test.

(g) Description

Figure D.28: Card Types for Ratio Test



Divisor

$a|b$

In mathematics, a divisor of an integer n, also called a factor of n, is an integer m that may be multiplied by some integer to produce n. In this case, one also says that n is a multiple of m.

(a) Description-Formula-Title

Divisor

$a|b$

(b) Formula-Title

Divisor

In mathematics, a divisor of an integer n, also called a factor of n, is an integer m that may be multiplied by some integer to produce n. In this case, one also says that n is a multiple of m.

(c) Description-Title

$a|b$

In mathematics, a divisor of an integer n, also called a factor of n, is an integer m that may be multiplied by some integer to produce n. In this case, one also says that n is a multiple of m.

(d) Description-Formula

$a|b$

(e) Formula

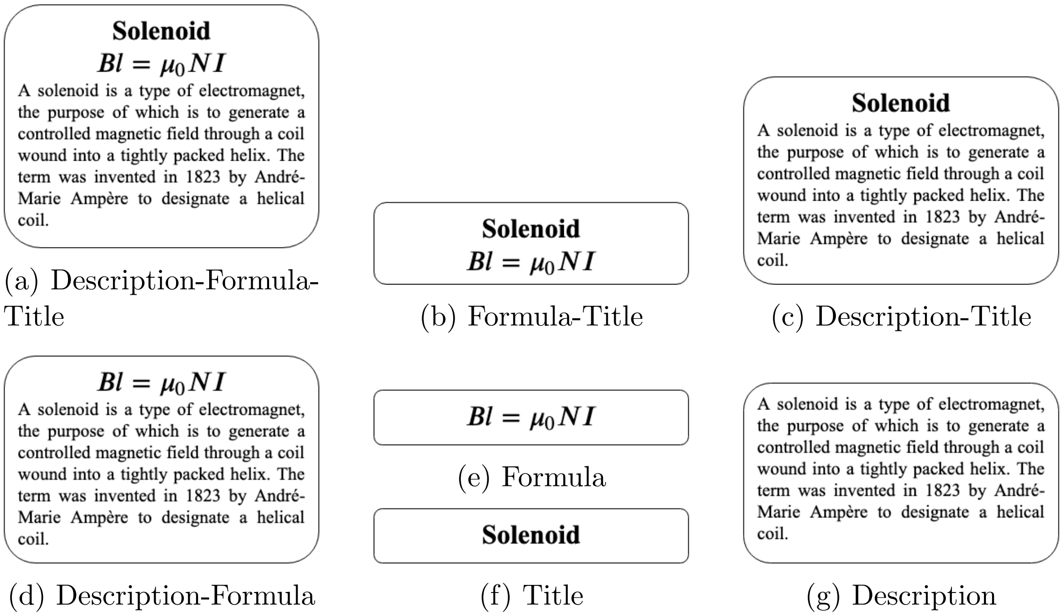
Divisor

(f) Title

In mathematics, a divisor of an integer n, also called a factor of n, is an integer m that may be multiplied by some integer to produce n. In this case, one also says that n is a multiple of m.

(g) Description

Figure D.29: Card Types for Divisor



Solenoid

$Bl = \mu_0 NI$

A solenoid is a type of electromagnet, the purpose of which is to generate a controlled magnetic field through a coil wound into a tightly packed helix. The term was invented in 1823 by André-Marie Ampère to designate a helical coil.

(a) Description-Formula-Title

Solenoid

$Bl = \mu_0 NI$

(b) Formula-Title

Solenoid

A solenoid is a type of electromagnet, the purpose of which is to generate a controlled magnetic field through a coil wound into a tightly packed helix. The term was invented in 1823 by André-Marie Ampère to designate a helical coil.

(c) Description-Title

$Bl = \mu_0 NI$

A solenoid is a type of electromagnet, the purpose of which is to generate a controlled magnetic field through a coil wound into a tightly packed helix. The term was invented in 1823 by André-Marie Ampère to designate a helical coil.

(d) Description-Formula

$Bl = \mu_0 NI$

(e) Formula

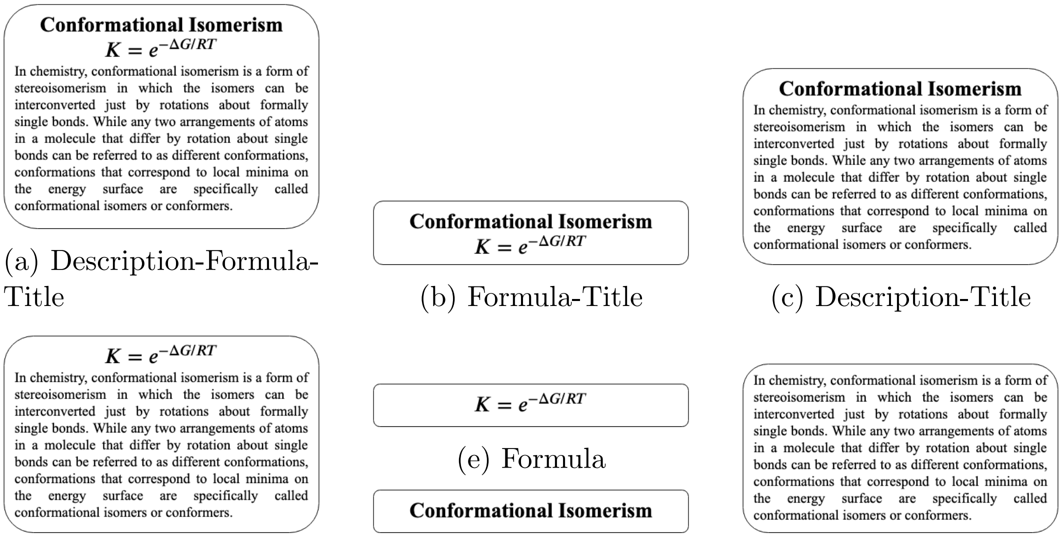
Solenoid

(f) Title

A solenoid is a type of electromagnet, the purpose of which is to generate a controlled magnetic field through a coil wound into a tightly packed helix. The term was invented in 1823 by André-Marie Ampère to designate a helical coil.

(g) Description

Figure D.30: Card Types for Solenoid



Conformational Isomerism

$K = e^{-\Delta G/RT}$

In chemistry, conformational isomerism is a form of stereoisomerism in which the isomers can be interconverted just by rotations about formally single bonds. While any two arrangements of atoms in a molecule that differ by rotation about single bonds can be referred to as different conformations, conformations that correspond to local minima on the energy surface are specifically called conformational isomers or conformers.

(a) Description-Formula-Title

Conformational Isomerism

$K = e^{-\Delta G/RT}$

(b) Formula-Title

Conformational Isomerism

In chemistry, conformational isomerism is a form of stereoisomerism in which the isomers can be interconverted just by rotations about formally single bonds. While any two arrangements of atoms in a molecule that differ by rotation about single bonds can be referred to as different conformations, conformations that correspond to local minima on the energy surface are specifically called conformational isomers or conformers.

(c) Description-Title

$K = e^{-\Delta G/RT}$

In chemistry, conformational isomerism is a form of stereoisomerism in which the isomers can be interconverted just by rotations about formally single bonds. While any two arrangements of atoms in a molecule that differ by rotation about single bonds can be referred to as different conformations, conformations that correspond to local minima on the energy surface are specifically called conformational isomers or conformers.

(d) Description-Formula

$K = e^{-\Delta G/RT}$

(e) Formula

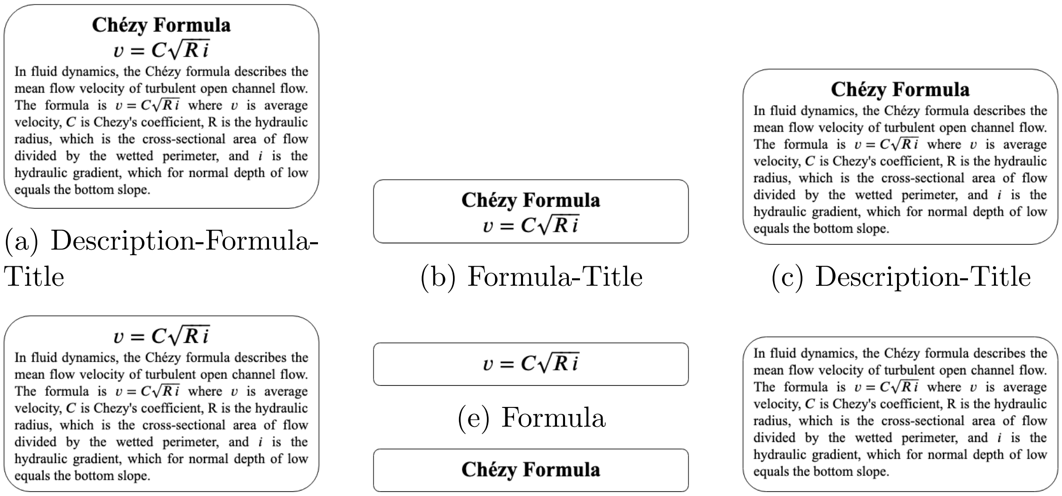
Conformational Isomerism

(f) Title

In chemistry, conformational isomerism is a form of stereoisomerism in which the isomers can be interconverted just by rotations about formally single bonds. While any two arrangements of atoms in a molecule that differ by rotation about single bonds can be referred to as different conformations, conformations that correspond to local minima on the energy surface are specifically called conformational isomers or conformers.

(g) Description

Figure D.31: Card Types for Conformational Isomerism



Chézy Formula

$v = C\sqrt{R\,i}$

In fluid dynamics, the Chézy formula describes the mean flow velocity of turbulent open channel flow. The formula is $v = C\sqrt{R\,i}$ where v is average velocity, C is Chezy's coefficient, R is the hydraulic radius, which is the cross-sectional area of flow divided by the wetted perimeter, and i is the hydraulic gradient, which for normal depth of low equals the bottom slope.

(a) Description-Formula-Title

Chézy Formula

$v = C\sqrt{R\,i}$

(b) Formula-Title

Chézy Formula

In fluid dynamics, the Chézy formula describes the mean flow velocity of turbulent open channel flow. The formula is $v = C\sqrt{R\,i}$ where v is average velocity, C is Chezy's coefficient, R is the hydraulic radius, which is the cross-sectional area of flow divided by the wetted perimeter, and i is the hydraulic gradient, which for normal depth of low equals the bottom slope.

(c) Description-Title

$v = C\sqrt{R\,i}$

In fluid dynamics, the Chézy formula describes the mean flow velocity of turbulent open channel flow. The formula is $v = C\sqrt{R\,i}$ where v is average velocity, C is Chezy's coefficient, R is the hydraulic radius, which is the cross-sectional area of flow divided by the wetted perimeter, and i is the hydraulic gradient, which for normal depth of low equals the bottom slope.

(d) Description-Formula

$v = C\sqrt{R\,i}$

(e) Formula

Chézy Formula

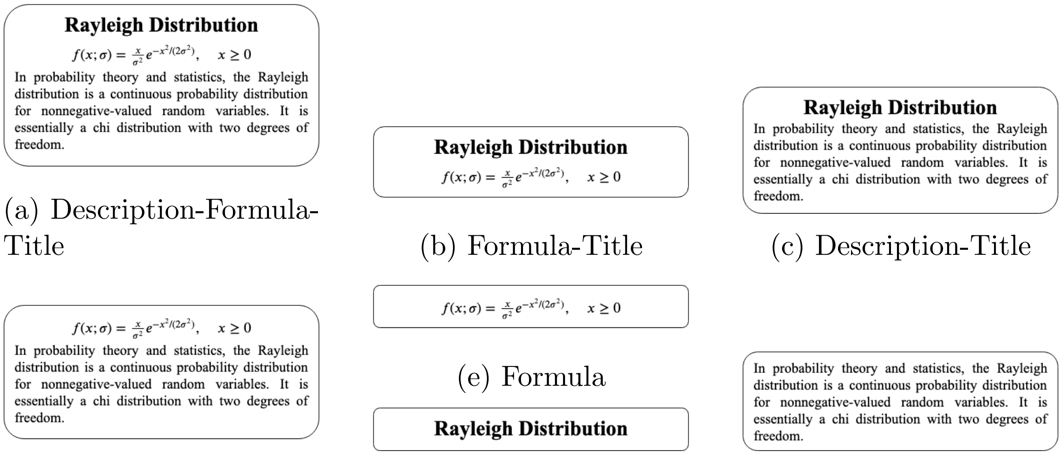
(f) Title

In fluid dynamics, the Chézy formula describes the mean flow velocity of turbulent open channel flow. The formula is $v = C\sqrt{R\,i}$ where v is average velocity, C is Chezy's coefficient, R is the hydraulic radius, which is the cross-sectional area of flow divided by the wetted perimeter, and i is the hydraulic gradient, which for normal depth of low equals the bottom slope.

(g) Description

Figure D.32: Card Types for Ch´zy Formula

D.3 Large Formulas



Rayleigh Distribution

$f(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0$

In probability theory and statistics, the Rayleigh distribution is a continuous probability distribution for nonnegative-valued random variables. It is essentially a chi distribution with two degrees of freedom.

(a) Description-Formula-Title

Rayleigh Distribution

$f(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0$

(b) Formula-Title

Rayleigh Distribution

In probability theory and statistics, the Rayleigh distribution is a continuous probability distribution for nonnegative-valued random variables. It is essentially a chi distribution with two degrees of freedom.

(c) Description-Title

$f(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0$

In probability theory and statistics, the Rayleigh distribution is a continuous probability distribution for nonnegative-valued random variables. It is essentially a chi distribution with two degrees of freedom.

(d) Description-Formula

$f(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0$

(e) Formula

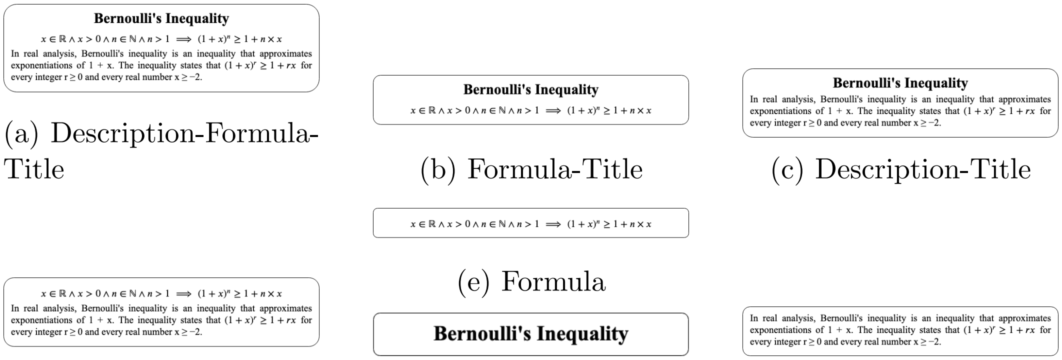
Rayleigh Distribution

(f) Title

In probability theory and statistics, the Rayleigh distribution is a continuous probability distribution for nonnegative-valued random variables. It is essentially a chi distribution with two degrees of freedom.

(g) Description

Figure D.33: Card Types for Rayleigh Distribution



Bernoulli's Inequality

$x \in \mathbb{R} \wedge x > 0 \wedge n \in \mathbb{N} \wedge n > 1 \implies (1+x)^n \geq 1 + n \times x$

In real analysis, Bernoulli's inequality is an inequality that approximates exponentiations of 1 + x. The inequality states that $(1+x)^r \geq 1 + rx$ for every integer r ≥ 0 and every real number x ≥ −2.

(a) Description-Formula-Title

Bernoulli's Inequality

$x \in \mathbb{R} \wedge x > 0 \wedge n \in \mathbb{N} \wedge n > 1 \implies (1+x)^n \geq 1 + n \times x$

(b) Formula-Title

Bernoulli's Inequality

In real analysis, Bernoulli's inequality is an inequality that approximates exponentiations of 1 + x. The inequality states that $(1+x)^r \geq 1 + rx$ for every integer r ≥ 0 and every real number x ≥ −2.

(c) Description-Title

$x \in \mathbb{R} \wedge x > 0 \wedge n \in \mathbb{N} \wedge n > 1 \implies (1+x)^n \geq 1 + n \times x$

In real analysis, Bernoulli's inequality is an inequality that approximates exponentiations of 1 + x. The inequality states that $(1+x)^r \geq 1 + rx$ for every integer r ≥ 0 and every real number x ≥ −2.

(d) Description-Formula

$x \in \mathbb{R} \wedge x > 0 \wedge n \in \mathbb{N} \wedge n > 1 \implies (1+x)^n \geq 1 + n \times x$

(e) Formula

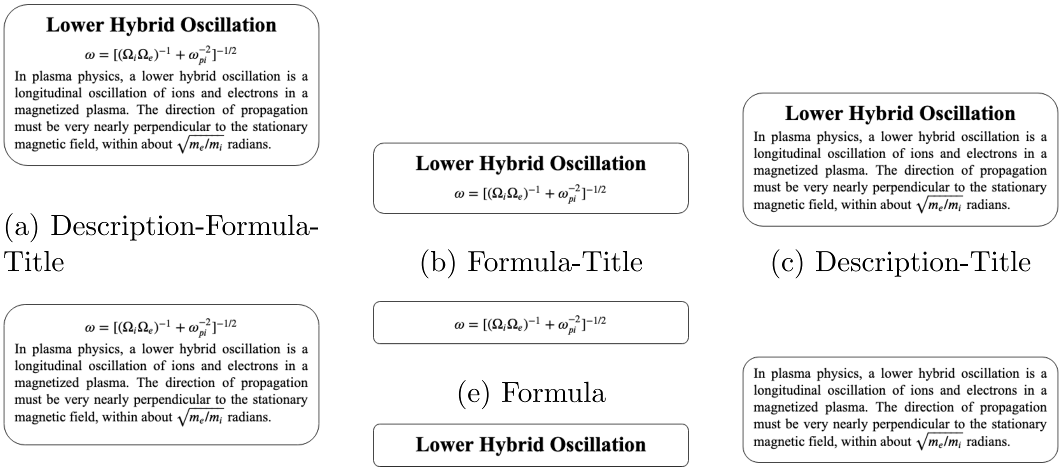
Bernoulli's Inequality

(f) Title

In real analysis, Bernoulli's inequality is an inequality that approximates exponentiations of 1 + x. The inequality states that $(1+x)^r \geq 1 + rx$ for every integer r ≥ 0 and every real number x ≥ −2.

(g) Description

Figure D.34: Card Types for Bernoulli's Inequality



Lower Hybrid Oscillation

$\omega = [(\Omega_i\Omega_e)^{-1} + \omega_{pi}^{-2}]^{-1/2}$

In plasma physics, a lower hybrid oscillation is a longitudinal oscillation of ions and electrons in a magnetized plasma. The direction of propagation must be very nearly perpendicular to the stationary magnetic field, within about $\sqrt{m_e/m_i}$ radians.

(a) Description-Formula-Title

Lower Hybrid Oscillation

$\omega = [(\Omega_i\Omega_e)^{-1} + \omega_{pi}^{-2}]^{-1/2}$

(b) Formula-Title

Lower Hybrid Oscillation

In plasma physics, a lower hybrid oscillation is a longitudinal oscillation of ions and electrons in a magnetized plasma. The direction of propagation must be very nearly perpendicular to the stationary magnetic field, within about $\sqrt{m_e/m_i}$ radians.

(c) Description-Title

$\omega = [(\Omega_i\Omega_e)^{-1} + \omega_{pi}^{-2}]^{-1/2}$

In plasma physics, a lower hybrid oscillation is a longitudinal oscillation of ions and electrons in a magnetized plasma. The direction of propagation must be very nearly perpendicular to the stationary magnetic field, within about $\sqrt{m_e/m_i}$ radians.

(d) Description-Formula

$\omega = [(\Omega_i\Omega_e)^{-1} + \omega_{pi}^{-2}]^{-1/2}$

(e) Formula

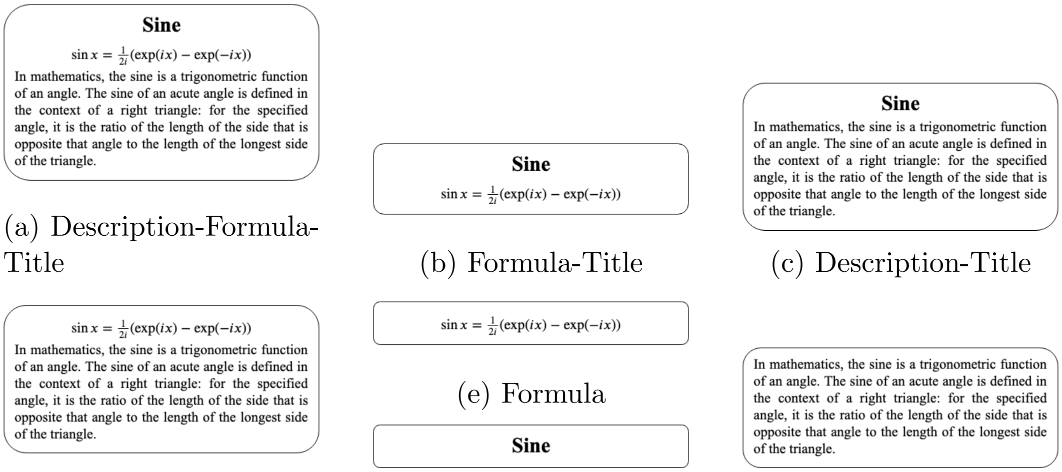
Lower Hybrid Oscillation

(f) Title

In plasma physics, a lower hybrid oscillation is a longitudinal oscillation of ions and electrons in a magnetized plasma. The direction of propagation must be very nearly perpendicular to the stationary magnetic field, within about $\sqrt{m_e/m_i}$ radians.

(g) Description

Figure D.35: Card Types for Lower Hybrid Oscillation



Sine

$\sin x = \frac{1}{2i}(\exp(ix) - \exp(-ix))$

In mathematics, the sine is a trigonometric function of an angle. The sine of an acute angle is defined in the context of a right triangle: for the specified angle, it is the ratio of the length of the side that is opposite that angle to the length of the longest side of the triangle.

(a) Description-Formula-Title

Sine

$\sin x = \frac{1}{2i}(\exp(ix) - \exp(-ix))$

(b) Formula-Title

Sine

In mathematics, the sine is a trigonometric function of an angle. The sine of an acute angle is defined in the context of a right triangle: for the specified angle, it is the ratio of the length of the side that is opposite that angle to the length of the longest side of the triangle.

(c) Description-Title

$\sin x = \frac{1}{2i}(\exp(ix) - \exp(-ix))$

In mathematics, the sine is a trigonometric function of an angle. The sine of an acute angle is defined in the context of a right triangle: for the specified angle, it is the ratio of the length of the side that is opposite that angle to the length of the longest side of the triangle.

(d) Description-Formula

$\sin x = \frac{1}{2i}(\exp(ix) - \exp(-ix))$

(e) Formula

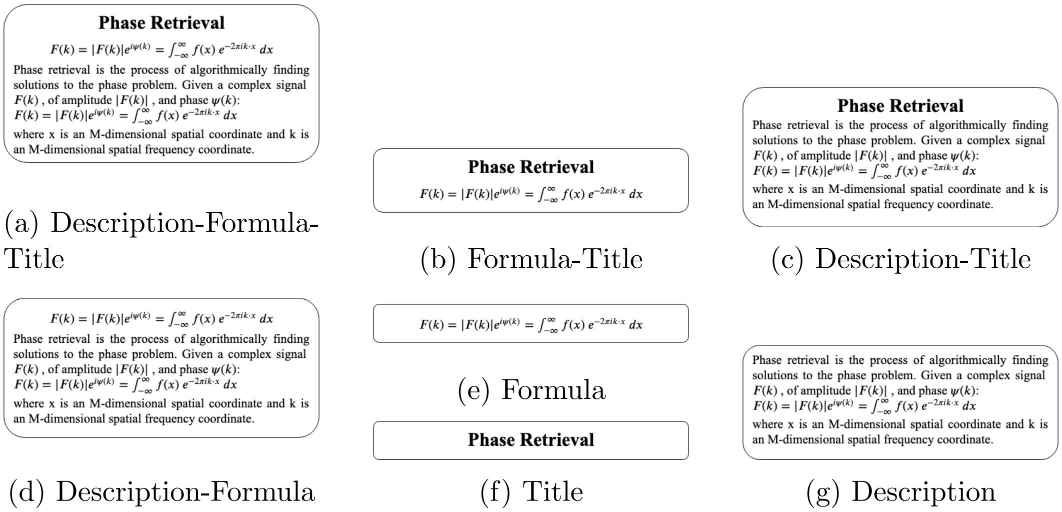
Sine

(f) Title

In mathematics, the sine is a trigonometric function of an angle. The sine of an acute angle is defined in the context of a right triangle: for the specified angle, it is the ratio of the length of the side that is opposite that angle to the length of the longest side of the triangle.

(g) Description

Figure D.36: Card Types for Sine



Phase Retrieval

$F(k) = |F(k)|e^{i\psi(k)} = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i k\cdot x}\, dx$

Phase retrieval is the process of algorithmically finding solutions to the phase problem. Given a complex signal $F(k)$, of amplitude $|F(k)|$, and phase $\psi(k)$:
$F(k) = |F(k)|e^{i\psi(k)} = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i k\cdot x}\, dx$
where x is an M-dimensional spatial coordinate and k is an M-dimensional spatial frequency coordinate.

(a) Description-Formula-Title

Phase Retrieval

$F(k) = |F(k)|e^{i\psi(k)} = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i k\cdot x}\, dx$

(b) Formula-Title

Phase Retrieval

Phase retrieval is the process of algorithmically finding solutions to the phase problem. Given a complex signal $F(k)$, of amplitude $|F(k)|$, and phase $\psi(k)$:
$F(k) = |F(k)|e^{i\psi(k)} = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i k\cdot x}\, dx$
where x is an M-dimensional spatial coordinate and k is an M-dimensional spatial frequency coordinate.

(c) Description-Title

$F(k) = |F(k)|e^{i\psi(k)} = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i k\cdot x}\, dx$

Phase retrieval is the process of algorithmically finding solutions to the phase problem. Given a complex signal $F(k)$, of amplitude $|F(k)|$, and phase $\psi(k)$:
$F(k) = |F(k)|e^{i\psi(k)} = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i k\cdot x}\, dx$
where x is an M-dimensional spatial coordinate and k is an M-dimensional spatial frequency coordinate.

(d) Description-Formula

$F(k) = |F(k)|e^{i\psi(k)} = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i k\cdot x}\, dx$

(e) Formula

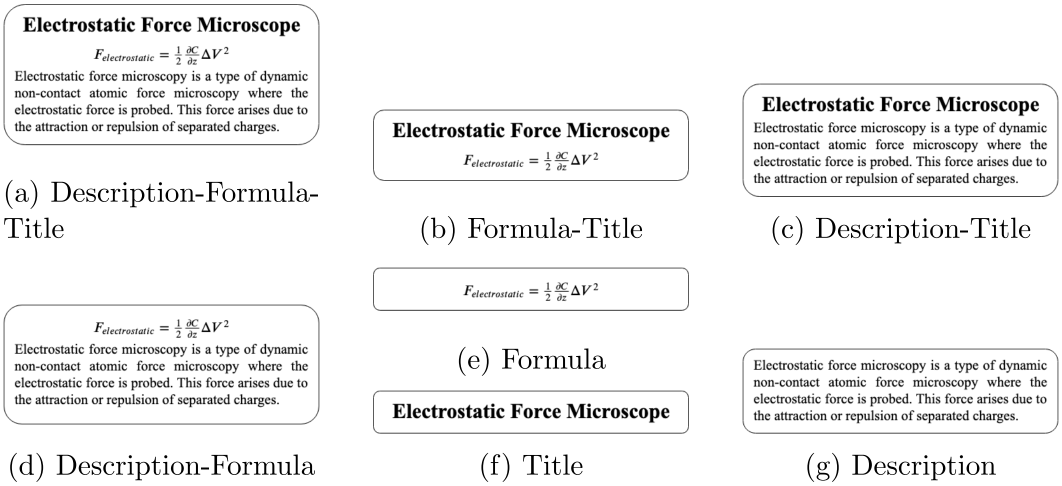
Phase Retrieval

(f) Title

Phase retrieval is the process of algorithmically finding solutions to the phase problem. Given a complex signal $F(k)$, of amplitude $|F(k)|$, and phase $\psi(k)$:
$F(k) = |F(k)|e^{i\psi(k)} = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i k\cdot x}\, dx$
where x is an M-dimensional spatial coordinate and k is an M-dimensional spatial frequency coordinate.

(g) Description

Figure D.37: Card Types for Phase Retrieval



Electrostatic Force Microscope

$F_{electrostatic} = \frac{1}{2}\frac{\partial C}{\partial z}\Delta V^2$

Electrostatic force microscopy is a type of dynamic non-contact atomic force microscopy where the electrostatic force is probed. This force arises due to the attraction or repulsion of separated charges.

(a) Description-Formula-Title

Electrostatic Force Microscope

$F_{electrostatic} = \frac{1}{2}\frac{\partial C}{\partial z}\Delta V^2$

(b) Formula-Title

Electrostatic Force Microscope

Electrostatic force microscopy is a type of dynamic non-contact atomic force microscopy where the electrostatic force is probed. This force arises due to the attraction or repulsion of separated charges.

(c) Description-Title

$F_{electrostatic} = \frac{1}{2}\frac{\partial C}{\partial z}\Delta V^2$

Electrostatic force microscopy is a type of dynamic non-contact atomic force microscopy where the electrostatic force is probed. This force arises due to the attraction or repulsion of separated charges.

(d) Description-Formula

$F_{electrostatic} = \frac{1}{2}\frac{\partial C}{\partial z}\Delta V^2$

(e) Formula

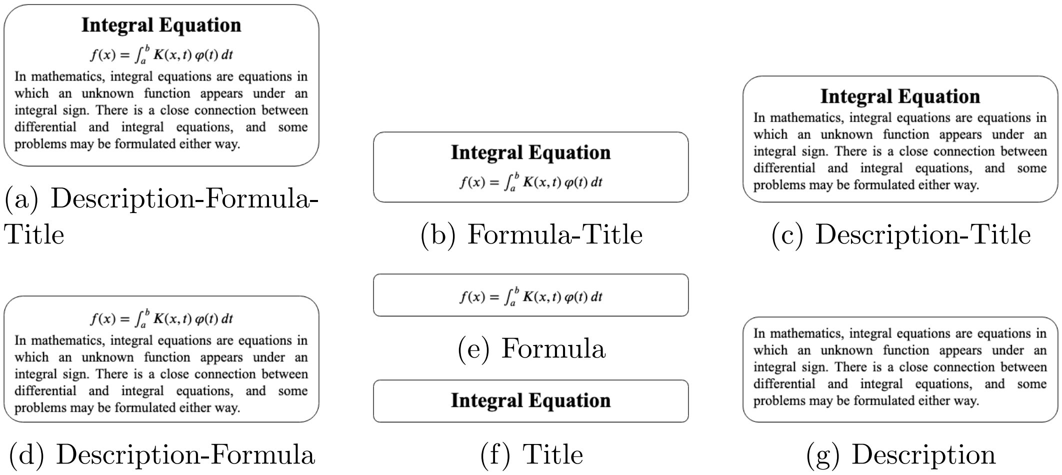
Electrostatic Force Microscope

(f) Title

Electrostatic force microscopy is a type of dynamic non-contact atomic force microscopy where the electrostatic force is probed. This force arises due to the attraction or repulsion of separated charges.

(g) Description

Figure D.38: Card Types for Electrostatic Force Microscope



Integral Equation

$f(x) = \int_a^b K(x,t)\,\varphi(t)\,dt$

In mathematics, integral equations are equations in which an unknown function appears under an integral sign. There is a close connection between differential and integral equations, and some problems may be formulated either way.

(a) Description-Formula-Title

Integral Equation

$f(x) = \int_a^b K(x,t)\,\varphi(t)\,dt$

(b) Formula-Title

Integral Equation

In mathematics, integral equations are equations in which an unknown function appears under an integral sign. There is a close connection between differential and integral equations, and some problems may be formulated either way.

(c) Description-Title

$f(x) = \int_a^b K(x,t)\,\varphi(t)\,dt$

In mathematics, integral equations are equations in which an unknown function appears under an integral sign. There is a close connection between differential and integral equations, and some problems may be formulated either way.

(d) Description-Formula

$f(x) = \int_a^b K(x,t)\,\varphi(t)\,dt$

(e) Formula

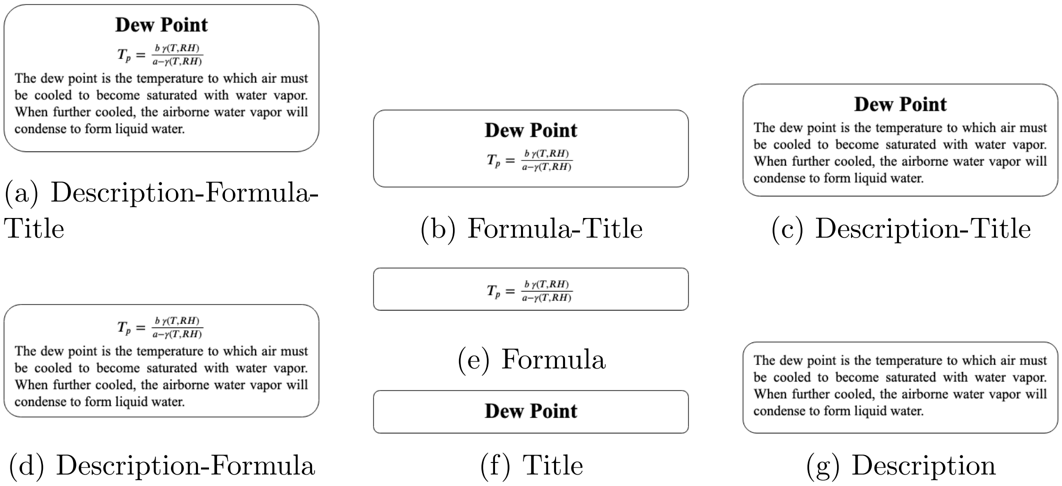
Integral Equation

(f) Title

In mathematics, integral equations are equations in which an unknown function appears under an integral sign. There is a close connection between differential and integral equations, and some problems may be formulated either way.

(g) Description

Figure D.39: Card Types for Integral Equation



Dew Point

$T_p = \frac{b\,\gamma(T,RH)}{a-\gamma(T,RH)}$

The dew point is the temperature to which air must be cooled to become saturated with water vapor. When further cooled, the airborne water vapor will condense to form liquid water.

(a) Description-Formula-Title

Dew Point

$T_p = \frac{b\,\gamma(T,RH)}{a-\gamma(T,RH)}$

(b) Formula-Title

Dew Point

The dew point is the temperature to which air must be cooled to become saturated with water vapor. When further cooled, the airborne water vapor will condense to form liquid water.

(c) Description-Title

$T_p = \frac{b\,\gamma(T,RH)}{a-\gamma(T,RH)}$

The dew point is the temperature to which air must be cooled to become saturated with water vapor. When further cooled, the airborne water vapor will condense to form liquid water.

(d) Description-Formula

$T_p = \frac{b\,\gamma(T,RH)}{a-\gamma(T,RH)}$

(e) Formula

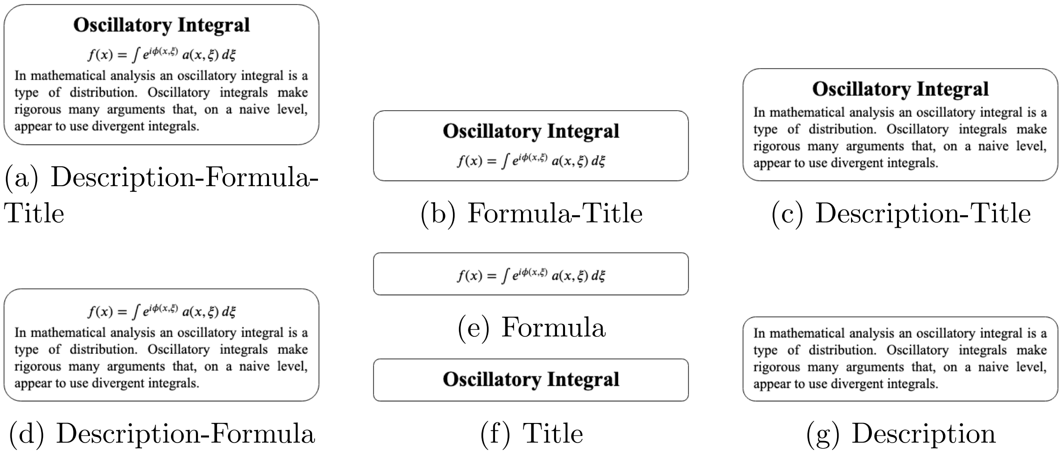
Dew Point

(f) Title

The dew point is the temperature to which air must be cooled to become saturated with water vapor. When further cooled, the airborne water vapor will condense to form liquid water.

(g) Description

Figure D.40: Card Types for Dew Point



Oscillatory Integral

$f(x) = \int e^{i\phi(x,\xi)}\, a(x,\xi)\, d\xi$

In mathematical analysis an oscillatory integral is a type of distribution. Oscillatory integrals make rigorous many arguments that, on a naive level, appear to use divergent integrals.

(a) Description-Formula-Title

Oscillatory Integral

$f(x) = \int e^{i\phi(x,\xi)}\, a(x,\xi)\, d\xi$

(b) Formula-Title

Oscillatory Integral

In mathematical analysis an oscillatory integral is a type of distribution. Oscillatory integrals make rigorous many arguments that, on a naive level, appear to use divergent integrals.

(c) Description-Title

$f(x) = \int e^{i\phi(x,\xi)}\, a(x,\xi)\, d\xi$

In mathematical analysis an oscillatory integral is a type of distribution. Oscillatory integrals make rigorous many arguments that, on a naive level, appear to use divergent integrals.

(d) Description-Formula

$f(x) = \int e^{i\phi(x,\xi)}\, a(x,\xi)\, d\xi$

(e) Formula

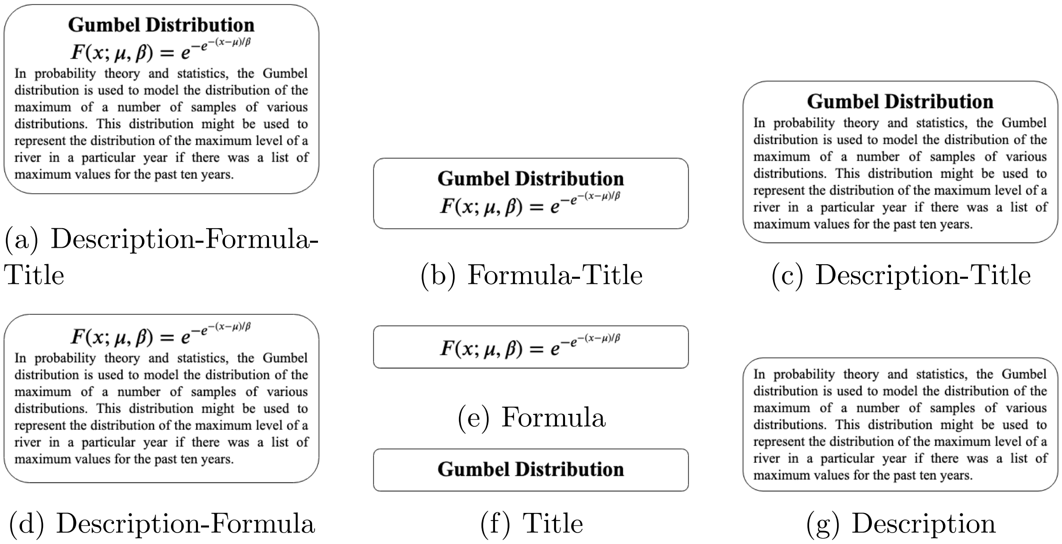
Oscillatory Integral

(f) Title

In mathematical analysis an oscillatory integral is a type of distribution. Oscillatory integrals make rigorous many arguments that, on a naive level, appear to use divergent integrals.

(g) Description

Figure D.41: Card Types for Oscillatory Integral



Gumbel Distribution

$F(x;\mu,\beta) = e^{-e^{-(x-\mu)/\beta}}$

In probability theory and statistics, the Gumbel distribution is used to model the distribution of the maximum of a number of samples of various distributions. This distribution might be used to represent the distribution of the maximum level of a river in a particular year if there was a list of maximum values for the past ten years.

(a) Description-Formula-Title

Gumbel Distribution

$F(x;\mu,\beta) = e^{-e^{-(x-\mu)/\beta}}$

(b) Formula-Title

Gumbel Distribution

In probability theory and statistics, the Gumbel distribution is used to model the distribution of the maximum of a number of samples of various distributions. This distribution might be used to represent the distribution of the maximum level of a river in a particular year if there was a list of maximum values for the past ten years.

(c) Description-Title

$F(x;\mu,\beta) = e^{-e^{-(x-\mu)/\beta}}$

In probability theory and statistics, the Gumbel distribution is used to model the distribution of the maximum of a number of samples of various distributions. This distribution might be used to represent the distribution of the maximum level of a river in a particular year if there was a list of maximum values for the past ten years.

(d) Description-Formula

$F(x;\mu,\beta) = e^{-e^{-(x-\mu)/\beta}}$

(e) Formula

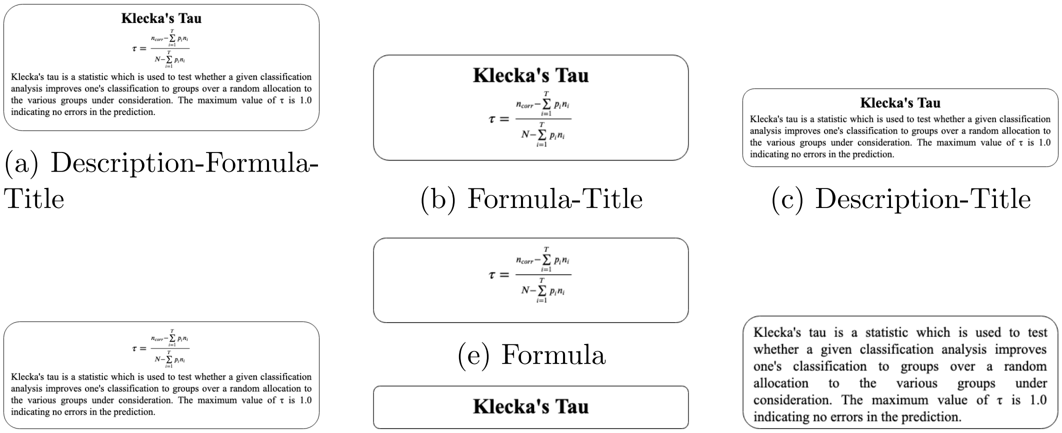
Gumbel Distribution

(f) Title

In probability theory and statistics, the Gumbel distribution is used to model the distribution of the maximum of a number of samples of various distributions. This distribution might be used to represent the distribution of the maximum level of a river in a particular year if there was a list of maximum values for the past ten years.

(g) Description

Figure D.42: Card Types for Gumbel Distribution



Klecka's Tau

$$\tau = \frac{n_{corr} - \sum_{i=1}^{T} p_i n_i}{N - \sum_{i=1}^{T} p_i n_i}$$

Klecka's tau is a statistic which is used to test whether a given classification analysis improves one's classification to groups over a random allocation to the various groups under consideration. The maximum value of τ is 1.0 indicating no errors in the prediction.

(a) Description-Formula-Title

Klecka's Tau

$$\tau = \frac{n_{corr} - \sum_{i=1}^{T} p_i n_i}{N - \sum_{i=1}^{T} p_i n_i}$$

(b) Formula-Title

Klecka's Tau

Klecka's tau is a statistic which is used to test whether a given classification analysis improves one's classification to groups over a random allocation to the various groups under consideration. The maximum value of τ is 1.0 indicating no errors in the prediction.

(c) Description-Title

$$\tau = \frac{n_{corr} - \sum_{i=1}^{T} p_i n_i}{N - \sum_{i=1}^{T} p_i n_i}$$

Klecka's tau is a statistic which is used to test whether a given classification analysis improves one's classification to groups over a random allocation to the various groups under consideration. The maximum value of τ is 1.0 indicating no errors in the prediction.

(d) Description-Formula

$$\tau = \frac{n_{corr} - \sum_{i=1}^{T} p_i n_i}{N - \sum_{i=1}^{T} p_i n_i}$$

(e) Formula

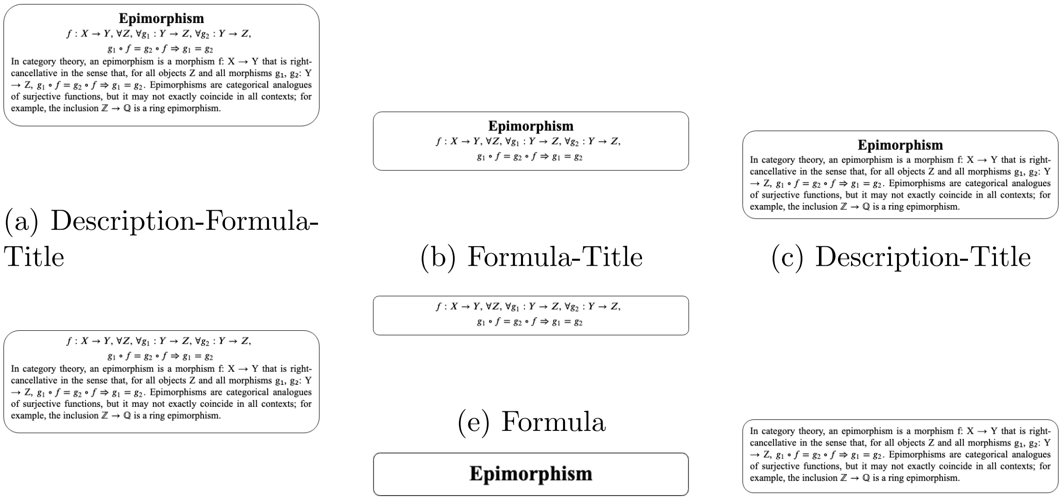
Klecka's Tau

(f) Title

Klecka's tau is a statistic which is used to test whether a given classification analysis improves one's classification to groups over a random allocation to the various groups under consideration. The maximum value of τ is 1.0 indicating no errors in the prediction.

(g) Description

Figure D.43: Card Types for Klecka's Tau



Epimorphism

$$f : X \to Y, \forall Z, \forall g_1 : Y \to Z, \forall g_2 : Y \to Z,$$
$$g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$$

In category theory, an epimorphism is a morphism f: X → Y that is right-cancellative in the sense that, for all objects Z and all morphisms g_1, g_2: Y → Z, $g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$. Epimorphisms are categorical analogues of surjective functions, but it may not exactly coincide in all contexts; for example, the inclusion ℤ → ℚ is a ring epimorphism.

(a) Description-Formula-Title

Epimorphism

$$f : X \to Y, \forall Z, \forall g_1 : Y \to Z, \forall g_2 : Y \to Z,$$
$$g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$$

(b) Formula-Title

Epimorphism

In category theory, an epimorphism is a morphism f: X → Y that is right-cancellative in the sense that, for all objects Z and all morphisms g_1, g_2: Y → Z, $g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$. Epimorphisms are categorical analogues of surjective functions, but it may not exactly coincide in all contexts; for example, the inclusion ℤ → ℚ is a ring epimorphism.

(c) Description-Title

$$f : X \to Y, \forall Z, \forall g_1 : Y \to Z, \forall g_2 : Y \to Z,$$
$$g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$$

In category theory, an epimorphism is a morphism f: X → Y that is right-cancellative in the sense that, for all objects Z and all morphisms g_1, g_2: Y → Z, $g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$. Epimorphisms are categorical analogues of surjective functions, but it may not exactly coincide in all contexts; for example, the inclusion ℤ → ℚ is a ring epimorphism.

(d) Description-Formula

$$f : X \to Y, \forall Z, \forall g_1 : Y \to Z, \forall g_2 : Y \to Z,$$
$$g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$$

(e) Formula

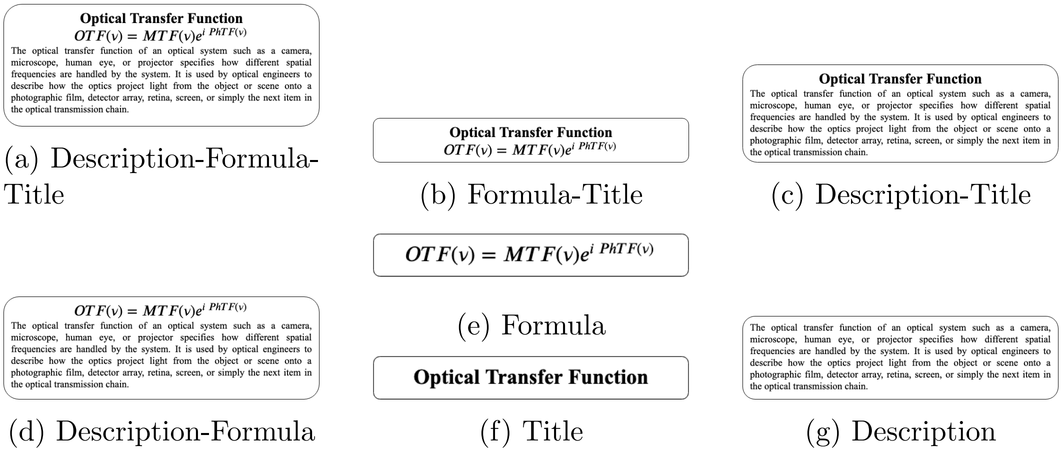
Epimorphism

(f) Title

In category theory, an epimorphism is a morphism f: X → Y that is right-cancellative in the sense that, for all objects Z and all morphisms g_1, g_2: Y → Z, $g_1 \circ f = g_2 \circ f \Rightarrow g_1 = g_2$. Epimorphisms are categorical analogues of surjective functions, but it may not exactly coincide in all contexts; for example, the inclusion ℤ → ℚ is a ring epimorphism.

(g) Description

Figure D.44: Card Types for Epimorphism



Optical Transfer Function

$OTF(v) = MTF(v)e^{i\ PhTF(v)}$

The optical transfer function of an optical system such as a camera, microscope, human eye, or projector specifies how different spatial frequencies are handled by the system. It is used by optical engineers to describe how the optics project light from the object or scene onto a photographic film, detector array, retina, screen, or simply the next item in the optical transmission chain.

(a) Description-Formula-Title

Optical Transfer Function

$OTF(v) = MTF(v)e^{i\ PhTF(v)}$

(b) Formula-Title

Optical Transfer Function

The optical transfer function of an optical system such as a camera, microscope, human eye, or projector specifies how different spatial frequencies are handled by the system. It is used by optical engineers to describe how the optics project light from the object or scene onto a photographic film, detector array, retina, screen, or simply the next item in the optical transmission chain.

(c) Description-Title

$OTF(v) = MTF(v)e^{i\ PhTF(v)}$

The optical transfer function of an optical system such as a camera, microscope, human eye, or projector specifies how different spatial frequencies are handled by the system. It is used by optical engineers to describe how the optics project light from the object or scene onto a photographic film, detector array, retina, screen, or simply the next item in the optical transmission chain.

(d) Description-Formula

$OTF(v) = MTF(v)e^{i\ PhTF(v)}$

(e) Formula

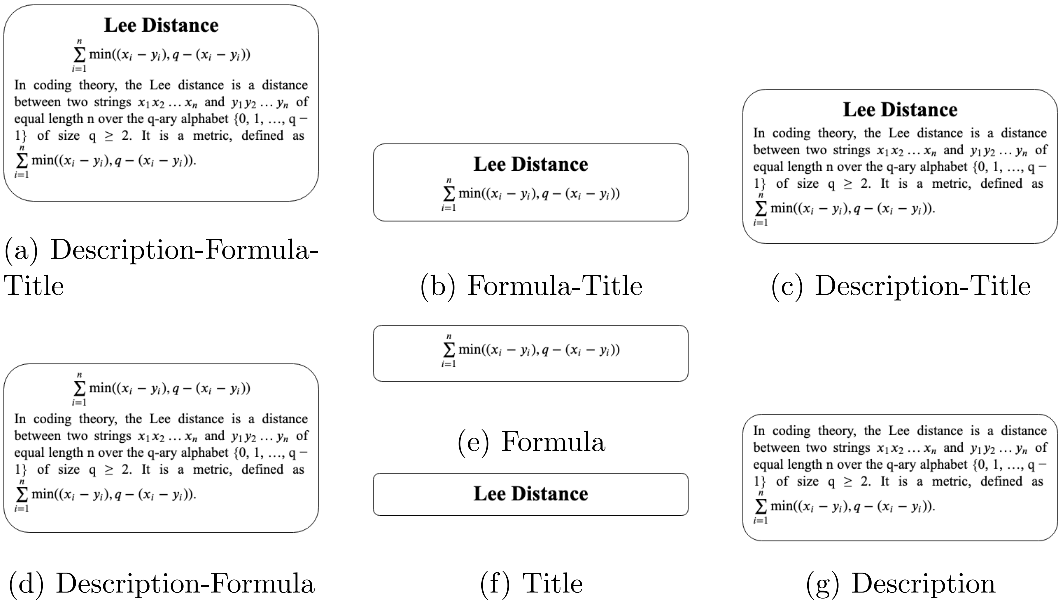
Optical Transfer Function

(f) Title

The optical transfer function of an optical system such as a camera, microscope, human eye, or projector specifies how different spatial frequencies are handled by the system. It is used by optical engineers to describe how the optics project light from the object or scene onto a photographic film, detector array, retina, screen, or simply the next item in the optical transmission chain.

(g) Description

Figure D.45: Card Types for Optical Transfer Function



Lee Distance

$\sum_{i=1}^{n} \min((x_i - y_i), q - (x_i - y_i))$

In coding theory, the Lee distance is a distance between two strings $x_1x_2\dots x_n$ and $y_1y_2\dots y_n$ of equal length n over the q-ary alphabet {0, 1, …, q − 1} of size q ≥ 2. It is a metric, defined as $\sum_{i=1}^{n} \min((x_i - y_i), q - (x_i - y_i))$.

(a) Description-Formula-Title

Lee Distance

$\sum_{i=1}^{n} \min((x_i - y_i), q - (x_i - y_i))$

(b) Formula-Title

Lee Distance

In coding theory, the Lee distance is a distance between two strings $x_1x_2\dots x_n$ and $y_1y_2\dots y_n$ of equal length n over the q-ary alphabet {0, 1, …, q − 1} of size q ≥ 2. It is a metric, defined as $\sum_{i=1}^{n} \min((x_i - y_i), q - (x_i - y_i))$.

(c) Description-Title

$\sum_{i=1}^{n} \min((x_i - y_i), q - (x_i - y_i))$

In coding theory, the Lee distance is a distance between two strings $x_1x_2\dots x_n$ and $y_1y_2\dots y_n$ of equal length n over the q-ary alphabet {0, 1, …, q − 1} of size q ≥ 2. It is a metric, defined as $\sum_{i=1}^{n} \min((x_i - y_i), q - (x_i - y_i))$.

(d) Description-Formula

$\sum_{i=1}^{n} \min((x_i - y_i), q - (x_i - y_i))$

(e) Formula

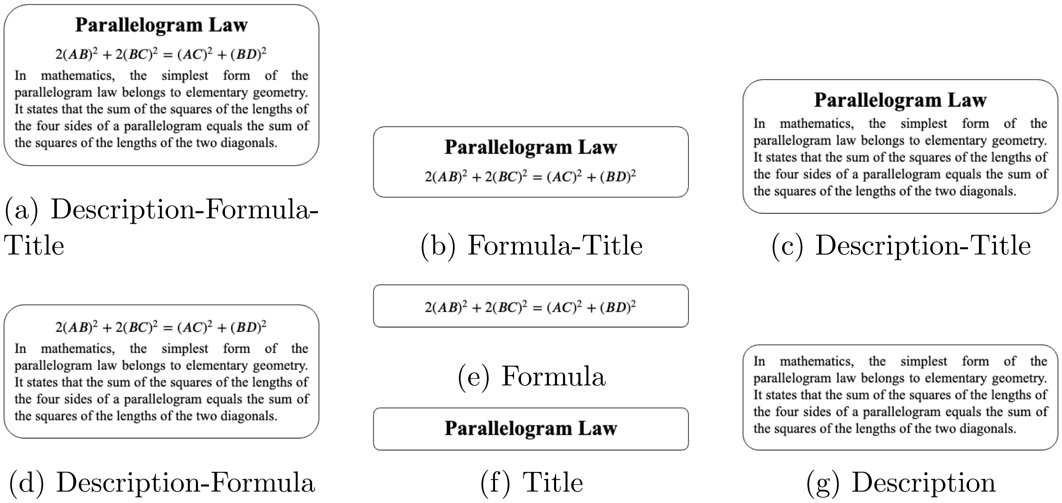
Lee Distance

(f) Title

In coding theory, the Lee distance is a distance between two strings $x_1x_2\dots x_n$ and $y_1y_2\dots y_n$ of equal length n over the q-ary alphabet {0, 1, …, q − 1} of size q ≥ 2. It is a metric, defined as $\sum_{i=1}^{n} \min((x_i - y_i), q - (x_i - y_i))$.

(g) Description

Figure D.46: Card Types for Lee Distance



Parallelogram Law

$$2(AB)^2 + 2(BC)^2 = (AC)^2 + (BD)^2$$

In mathematics, the simplest form of the parallelogram law belongs to elementary geometry. It states that the sum of the squares of the lengths of the four sides of a parallelogram equals the sum of the squares of the lengths of the two diagonals.

(a) Description-Formula-Title

Parallelogram Law

$$2(AB)^2 + 2(BC)^2 = (AC)^2 + (BD)^2$$

(b) Formula-Title

Parallelogram Law

In mathematics, the simplest form of the parallelogram law belongs to elementary geometry. It states that the sum of the squares of the lengths of the four sides of a parallelogram equals the sum of the squares of the lengths of the two diagonals.

(c) Description-Title

$$2(AB)^2 + 2(BC)^2 = (AC)^2 + (BD)^2$$

In mathematics, the simplest form of the parallelogram law belongs to elementary geometry. It states that the sum of the squares of the lengths of the four sides of a parallelogram equals the sum of the squares of the lengths of the two diagonals.

(d) Description-Formula

$$2(AB)^2 + 2(BC)^2 = (AC)^2 + (BD)^2$$

(e) Formula

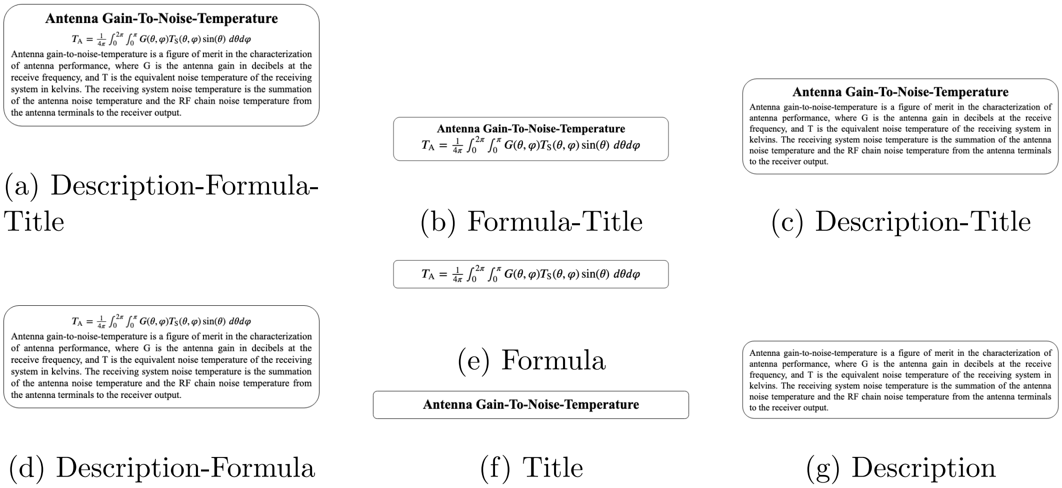
Parallelogram Law

(f) Title

In mathematics, the simplest form of the parallelogram law belongs to elementary geometry. It states that the sum of the squares of the lengths of the four sides of a parallelogram equals the sum of the squares of the lengths of the two diagonals.

(g) Description

Figure D.47: Card Types for Parallelogram Law



Antenna Gain-To-Noise-Temperature

$$T_\text{A} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} G(\theta,\varphi)T_\text{S}(\theta,\varphi)\sin(\theta)\ d\theta d\varphi$$

Antenna gain-to-noise-temperature is a figure of merit in the characterization of antenna performance, where G is the antenna gain in decibels at the receive frequency, and T is the equivalent noise temperature of the receiving system in kelvins. The receiving system noise temperature is the summation of the antenna noise temperature and the RF chain noise temperature from the antenna terminals to the receiver output.

(a) Description-Formula-Title

Antenna Gain-To-Noise-Temperature

$$T_\text{A} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} G(\theta,\varphi)T_\text{S}(\theta,\varphi)\sin(\theta)\ d\theta d\varphi$$

(b) Formula-Title

Antenna Gain-To-Noise-Temperature

Antenna gain-to-noise-temperature is a figure of merit in the characterization of antenna performance, where G is the antenna gain in decibels at the receive frequency, and T is the equivalent noise temperature of the receiving system in kelvins. The receiving system noise temperature is the summation of the antenna noise temperature and the RF chain noise temperature from the antenna terminals to the receiver output.

(c) Description-Title

$$T_\text{A} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} G(\theta,\varphi)T_\text{S}(\theta,\varphi)\sin(\theta)\ d\theta d\varphi$$

Antenna gain-to-noise-temperature is a figure of merit in the characterization of antenna performance, where G is the antenna gain in decibels at the receive frequency, and T is the equivalent noise temperature of the receiving system in kelvins. The receiving system noise temperature is the summation of the antenna noise temperature and the RF chain noise temperature from the antenna terminals to the receiver output.

(d) Description-Formula

$$T_\text{A} = \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} G(\theta,\varphi)T_\text{S}(\theta,\varphi)\sin(\theta)\ d\theta d\varphi$$

(e) Formula

Antenna Gain-To-Noise-Temperature

(f) Title

Antenna gain-to-noise-temperature is a figure of merit in the characterization of antenna performance, where G is the antenna gain in decibels at the receive frequency, and T is the equivalent noise temperature of the receiving system in kelvins. The receiving system noise temperature is the summation of the antenna noise temperature and the RF chain noise temperature from the antenna terminals to the receiver output.

(g) Description

Figure D.48: Card Types for Antenna Gain To Noise Temperature

Appendix E

Mathematical Concepts as per familiarity in Human Experiment

E.1 Symbols

Table E.1: Familiar and Less Familiar, Symbols used for Human Experiment. *Indicates Practice Trials

Sr No	Familiar	Less Familiar
1	Congruence	Aleph Number
2	Inequality	Converse Implication
3	Line Integral	Projective Space
4	Complex Conjugate	Compact Embedding
5	Cross Product	Entailment
6	Partial Derivative	Beth number
7	Plus-Minus	Wreath Product
8	Left-Open Interval	Covering Relation
9	Addition*	Semijoin*

E.2 Small Formulas

Table E.2: Familiar and Less Familiar, Small Formulas used for Human Experiment. *Indicates Practice Trials

Sr No	Familiar	Less Familiar
1	Adsorption	Autonomous Consumption
2	Rotating Unbalance	Classification of Electromagentic Fields
3	Magnetic Energy	Reality Structure
4	Mired	Allan Variance
5	Angular Velocity	Equianharmonic
6	Ratio Test	Hugh Cardinal
7	Divisor	Conformational Isomerism
8	Solenoid	Chézy Formula
9	Pythagorean Theorem*	

E.3 Large Formulas

Table E.3: Familiar and Less Familiar, Large Formulas used for Human Experiment. *Indicates Practice Trials

Sr No	Familiar	Less Familiar
1	Rayleigh Distribution	Lower Hybrid Oscillation
2	Bernoulli's Inequality	Phase Retrieval
3	Sine	Electrostatic Force Microscope
4	Integral Equation	Oscillatory Integral
5	Dew Point	Gumbel Distribution
6	Optical Transfer Function	Klecka's Tau
7	Parallelogram Law	Epimorphism
8	Antenna Gain To Noise Temperature	Lee Distance
9	Differntial Entropy*	

Appendix F

Secondary Results

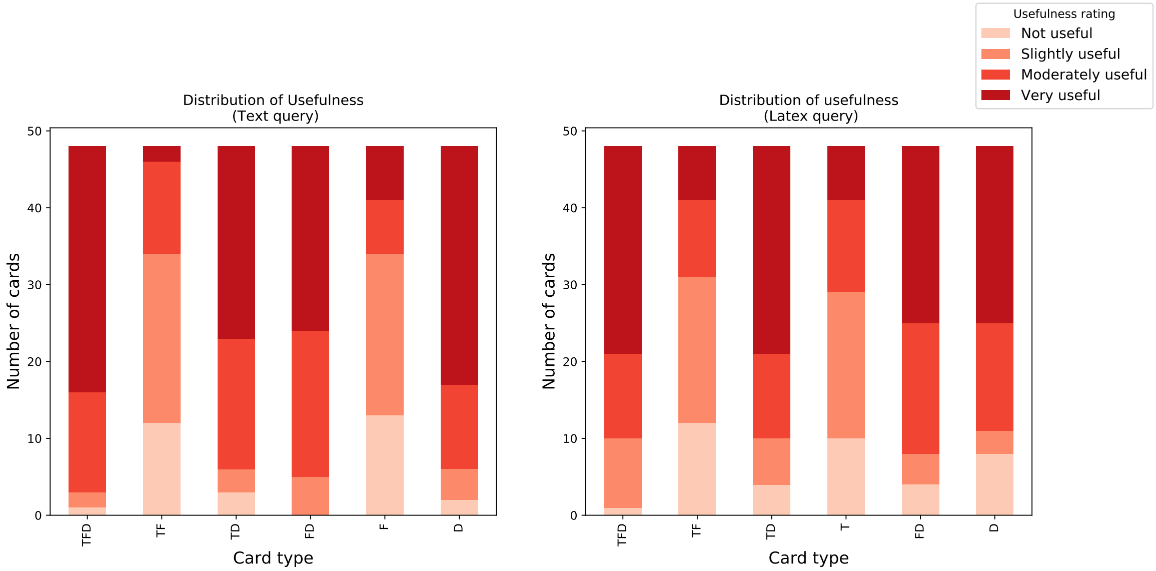


Figure F.1: Distribution of Usefulness Scores across total number of cards for Familiar Concept

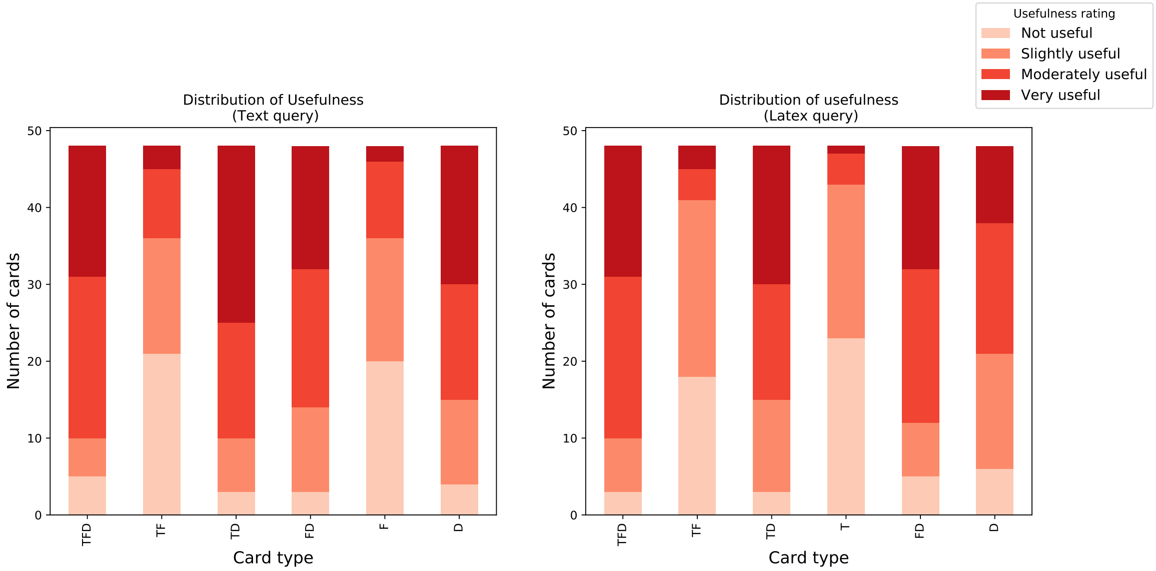


Figure F.2: Distribution of Usefulness Scores across total number of cards for Less Familiar Concept

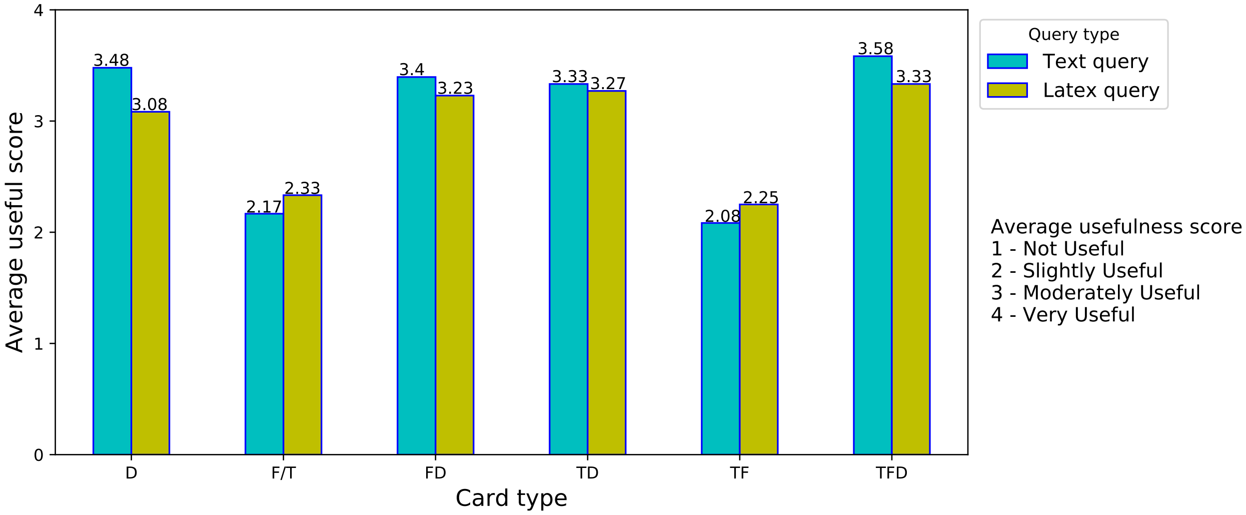


Figure F.3: Average Usefulness Scores per Card Type for Familiar Concepts

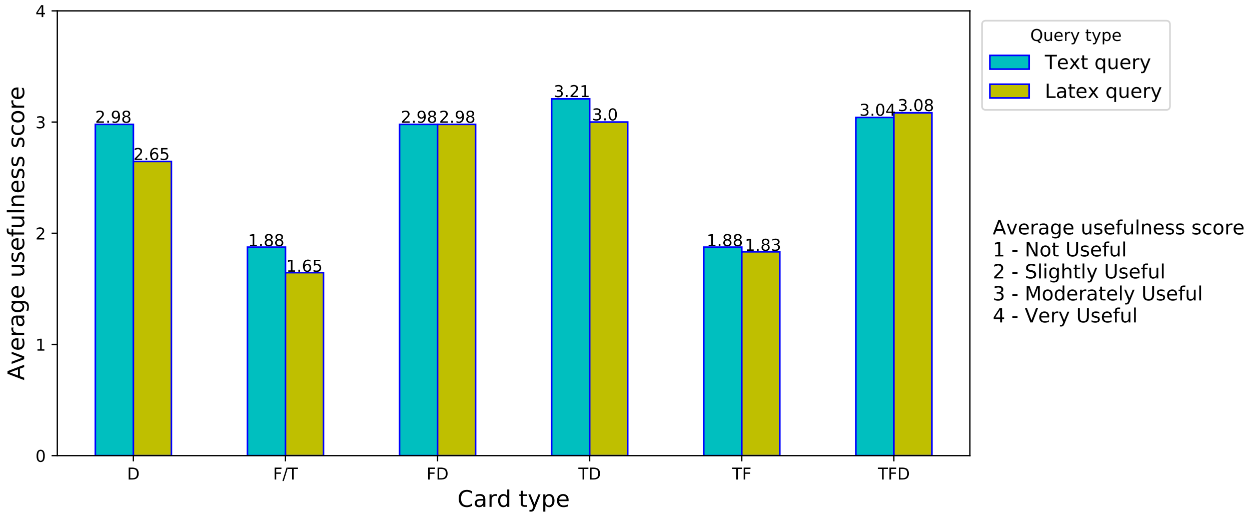


Figure F.4: Average Usefulness Scores per Card Type for Less Familiar Concepts

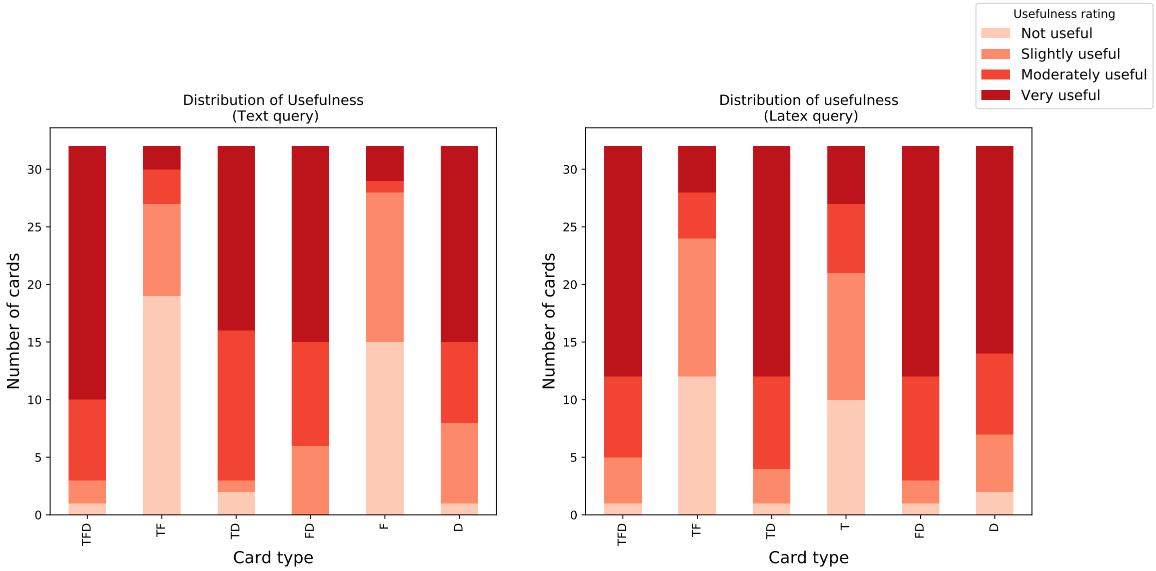


Figure F.5: Distribution of Usefulness Scores across total number of cards for Symbols

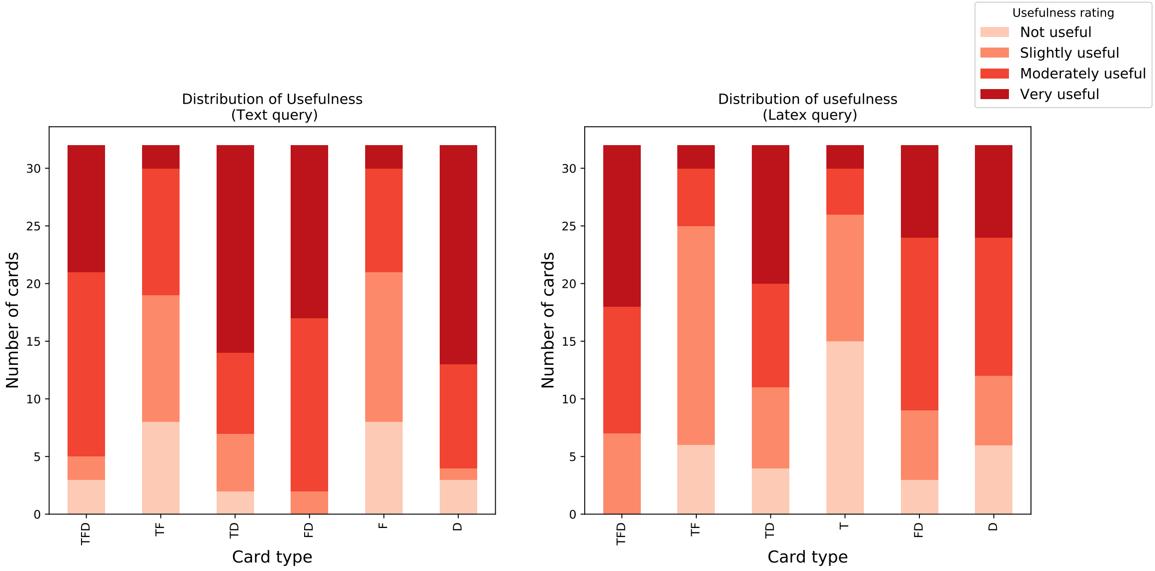


Figure F.6: Distribution of Usefulness Scores across total number of cards for Small Formulas

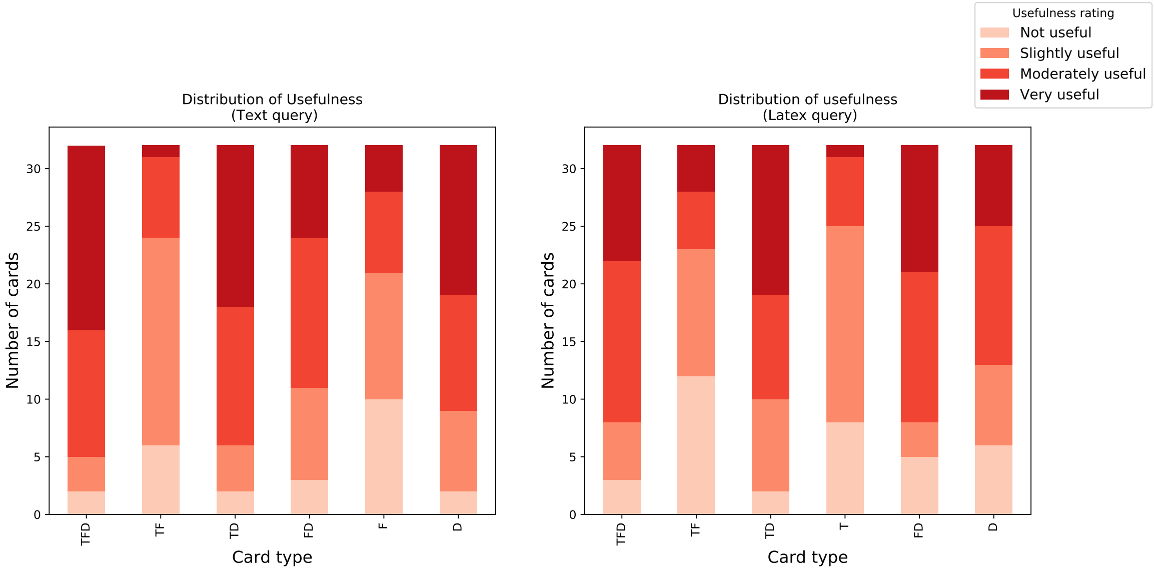


Figure F.7: Distribution of Usefulness Scores across total number of cards for Large Formulas

Bibliography

[1] Ron Ausbrooks, Stephen Buswell, David Carlisle, Giorgi Chavchanidze, Stphane Dalmas, Stan Devitt, Angel Diaz, Sam Dooley, Roger Hunter, Patrick Ion, Michael Kohlhase, Azzeddine Lazrek, Paul Libbrecht, Bruce Miller, Robert Miner, Murray Sargent, Bruce Smith, Neil Soiffer, Robert Sutor, and Stephen Watt. Mathematical markup language (mathml) version 3.0, 01 2010.

[2] Krisztian Balog. *Entity-Oriented Search*, volume 39 of *The Information Retrieval Series*. Springer, 2018.

[3] Horatiu Bota, Ke Zhou, and Joemon M. Jose. Playing your cards right: The effect of entity cards on search behaviour and workload. In *Proceedings of the 2016 ACM Conference on Human Information Interaction and Retrieval, CHIIR 2016, Carrboro, North Carolina, USA, March 13-17, 2016*, pages 131–140, 2016.

[4] Andrei Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.

[5] Stephen Buswell, Olga Caprotti, David Carlisle, Michael Dewar, Marc Gatano, and Michael Kohlhase. The open math standard, version 2.0, 01 2004.

[6] Kenny Davila and Richard Zanibbi. Layout and semantics: Combining representations for mathematical formula search. In *SIGIR*, pages 1165–1168. ACM, 2017.

[7] Andreas Franke and Michael Kohlhase. System description: Mbase, an open mathematical knowledge base. In *Automated Deduction - CADE-17, 17th International Conference on Automated Deduction, Pittsburgh, PA, USA, June 17-20, 2000, Proceedings*, pages 455–459, 2000.

[8] Kristianto Giovanni, Nghiem Minh, Matsubayashi Yuichiroh, and Aizawa Akiko. Extracting definitions of mathematical expressions in scientific papers. In *Proceedings of the 26th Annual Conference of JSAI. 2012.*, 2012.

[9] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. Dynamic factual summaries for entity cards. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, pages 773–782, 2017.

[10] Marti A. Hearst. *SEARCH USER INTERFACES*, chapter The Evaluation of Search User Interfaces. CAMBRIDGE UNIV PRESS, 2009.

[11] Xuan Hu, Liangcai Gao, Xiaoyan Lin, Zhi Tang, Xiaofan Lin, and Josef B. Baker. Wikimirs: a mathematical information retrieval system for wikipedia. In *JCDL*, pages 11–20. ACM, 2013.

[12] Mihnea Iancu, Michael Kohlhase, Florian Rabe, and Josef Urban. The mizar mathematical library in omdoc: Translation and applications. *J. Autom. Reasoning*, 50(2):191–202, 2013.

[13] Zhuoren Jiang, Liangcai Gao, Ke Yuan, Zheng Gao, Zhi Tang, and Xiaozhong Liu. Mathematics content understanding for cyberlearning via formula evolution map. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, pages 37–46, 2018.

[14] Jimmy, Guido Zuccon, Bevan Koopman, and Gianluca Demartini. Health cards for consumer health search. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019.*, pages 35–44, 2019.

[15] Gianluca Demartini Jimmy Guido Zuccon and Bevan Koopman. Health cards to assist decision making in consumer health search. In *AMIA 2019, American Medical Informatics Association Annual Symposium, Washington, D.C., November 16-20, 2019*, 2019.

[16] Michael Kohlhase. Omdoc: An open markup format for mathematical documents (version 1.1), 01 2001.

[17] Giovanni Yoko Kristianto and Akiko Aizawa. Linking mathematical expressions to Wikipedia. In *Proceedings of the 1st Workshop on Scholarly Web Mining*, SWM '17, pages 57–64, New York, NY, USA, 2017. ACM.

[18] Giovanni Yoko Kristianto, Goran Topic, and Akiko Aizawa. Exploiting textual descriptions and dependency graph for searching mathematical expressions in scientific papers. In *Ninth International Conference on Digital Information Management, ICDIM 2014, Phitsanulok, Thailand, September 29 - Oct. 1, 2014*, pages 110–117, 2014.

[19] Giovanni Yoko Kristianto, Goran Topic, and Akiko Aizawa. Extracting textual descriptions of mathematical expressions in scientific papers. *D-Lib Magazine*, 20(11/12), 2014.

[20] Giovanni Yoko Kristianto, Goran Topic, Minh-Quoc Nghiem, and Akiko N. Aizawa. Annotating scientific papers for mathematical formula search. In *Proceedings of the Fifth workshop on Exploiting Semantic Annotations in Information Retrieval, ESAIR 2012, Maui, HI, USA, October 28, 2012*, pages 17–18, 2012.

[21] Carla Teixeira Lopes and Hugo Sousa. Assisting health consumers while searching the web through medical annotations. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval, CHIIR 2019, Glasgow, Scotland, UK, March 10-14, 2019*, pages 219–223, 2019.

[22] Behrooz Mansouri, Shaurya Rohatgi, Douglas W. Oard, Jian Wu, C. Lee Giles, and Richard Zanibbi. Tangent-cft: An embedding model for mathematical formulas. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2019, Santa Clara, CA, USA, October 2-5, 2019.*, pages 11–18, 2019.

[23] Behrooz Mansouri, Richard Zanibbi, and Douglas W. Oard. Characterizing searches for mathematical concepts. In *19th ACM/IEEE Joint Conference on Digital Libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019*, pages 57–66, 2019.

[24] Olga Nevzorova, Nikita Zhiltsov, Danila Zaikin, Olga Zhibrik, Alexander Kirillovich, Vladimir Nevzorov, and Evgeniy Birialtsev. Bringing math to LOD: A semantic publishing platform prototype for scientific collections in mathematics. In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I*, pages 379–394, 2013.

[25] Minh Nghiem Quoc, Keisuke Yokoi, Yuichiroh Matsubayashi, and Akiko Aizawa. Mining coreference relations between formulas and text using wikipedia. In *Proceedings of the Second Workshop on NLP Challenges in the Information Explosion Era (NLPIX 2010)*, pages 69–74, Beijing, China, August 2010. Coling 2010 Organizing Committee.

[26] Nidhin Pattaniyil and Richard Zanibbi. Combining TF-IDF text retrieval with an inverted index over symbol pairs in math expressions: The tangent math search engine at NTCIR 2014. In *NTCIR*. National Institute of Informatics (NII), 2014.

[27] Sini Govinda Pillai, Lay-Ki Soon, and Su-Cheng Haw. Comparing dbpedia, wikidata, and yago for web information retrieval. In *Intelligent and Interactive Computing*, pages 525–535. Springer, 2019.

[28] Minh Nghiem Quoc, Keisuke Yokoi, Yuichiroh Matsubayashi, and Akiko Aizawa. Mining coreference relations between formulas and text using wikipedia. In *Proceedings of the Second Workshop on NLP Challenges in the Information Explosion Era (NLPIX 2010)*, pages 69–74, 2010.

[29] Matthias S. Reichenbach, Anurag Agarwal, and Richard Zanibbi. Rendering expressions to improve accuracy of relevance assessment for math search. In *The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Gold Coast , QLD, Australia - July 06 - 11, 2014*, pages 851–854, 2014.

[30] Remigiusz Sapa, Monika Krakowska, and Malgorzata Janiak. Information seeking behaviour of mathematicians: scientists and students. *Inf. Res.*, 19(4), 2014.

[31] Christopher Sasarak, Kevin Hart, Richard Pospesel, David Stalnaker, Lei Hu, Robert Livolsi, Siyu Zhu, and Richard Zanibbi. min: A multimodal web interface for math search. In *Symp. Human-Computer Interaction and Information Retrieval, Cambridge, MA*, 2012.

[32] Yiannos Stathopoulos and Simone Teufel. Mathematical information retrieval based on type embeddings and query expansion. In *COLING*, pages 2344–2355. ACL, 2016.

[33] Keita Del Valle Wangari, Richard Zanibbi, and Anurag Agarwal. Discovering real-world use cases for a multimodal math search interface. In *SIGIR*, pages 947–950. ACM, 2014.

[34] Keisuke Yokoi, Minh-Quoc Nghiem, Yuichiroh Matsubayashi, and Akiko Aizawa. Contextual analysis of mathematical expressions for advanced mathematical search. *Polibits*, 43:81–86, 2011.

[35] Richard Zanibbi and Awelemdy Orakwue. Math search for the masses: Multimodal search interfaces and appearance-based retrieval. In *CICM*, volume 9150 of *Lecture Notes in Computer Science*, pages 18–36. Springer, 2015.

[36] Jin Zhao, Min-Yen Kan, and Yin Leng Theng. Math information retrieval: user requirements and prototype implementation. In *JCDL*, pages 187–196. ACM, 2008.

[37] Wei Zhong and Richard Zanibbi. Structural similarity search for formulas using leaf-root paths in operator subtrees. In *Advances in Information Retrieval - 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14-18, 2019, Proceedings, Part I*, pages 116–129, 2019.