Introduction & Background
Advent of Web 2.0 popularized the importance of mashup. Today more than 18000 APIs are present on web. Discovery of relevant APIs is challenging. This project proposes to improvise recommendation of relevant APIs for mashup creation based on:
- API functionality.
- API’s usage history.
- API popularity.

Mashup: Combination of web applications.

Proposed Approach

Traditional Topic Modelling: LDA
- LDA (Latent Dirichlet Allocation) is a statistical topic modelling algorithm that works on Dirichlet distribution.
- It finds latent topics from the documents.
- Topics are distribution over a vocabulary.
- Document is a distribution over topics.

Collaborative Filtering
- Collaborative filtering is a well-known recommendation algorithm that works on database of preferences for items by users. It may be:
  - Memory Based
  - Model based

Matrix Factorization: Model Based
R = M x N
M, N decomposed into D dimension of latent preferences.
R predicted = (M x D) x (N x D)

Advanced Topic Modelling: HDP
- Hierarchical Dirichlet Process is a non-parametric Bayesian clustering algorithm.
- It clusters data possessing mixture model.
- Works on LDA + Gibbs Sampling.

Probabilistic Matrix Factorization
- Probabilistic approach.
- Works on Bayesian Perspective.
- Uses Gaussian distribution.

Combining Topic Modelling & Collaborative Filtering
- Probability of using an API given a mashup m is given by Bayes Theorem
  \[ p(a_1,a_2|m) p(m) = p(a_1|m) p(a_2|m) p(m) \]

Top k APIs selected based on popularity.