# Recommender System for Citations and Illustrations

Mandar Mitra

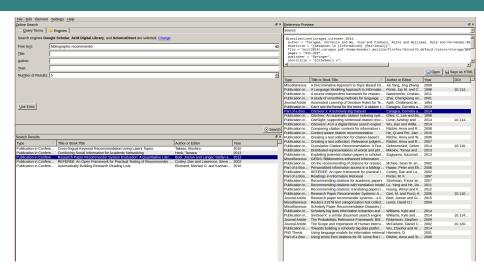
Indian Statistical Institute

# Motivation / problem statement

Create tool to assist an author who is writing paper / creating presentation

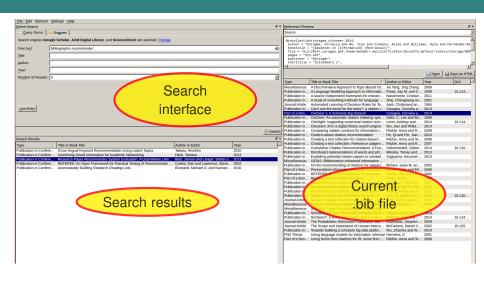
- Provide bibliographic details of paper to be cited tools exist
- Suggest papers to cite techniques exist
- Suggest appropriate illustrations / images for presentation exploratory

#### What we have



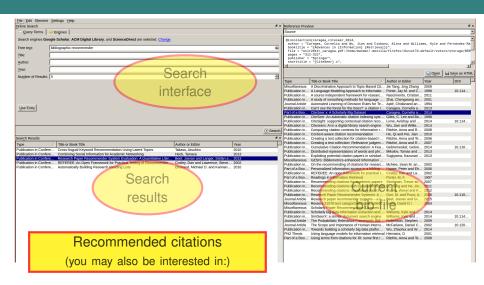
M. Mitra (ISI) CVPR-20/RSCI/MM 3

#### What we have



M. Mitra (ISI) CVPR-20/RSCI/MM 3

#### What we want



M. Mitra (ISI) CVPR-20/RSCI/MM 4

# Proposed work plan

- Year I: Literature survey (update to BEEL ET AL., 2016)
  - Select methods to compare empirically
  - Expand existing dataset
- Year II: Implementation and evaluation of methods identified above
  - Design of editor-agnostic recommendation service
- Year III: Implementation of service
  - Creation of prototype plugin
  - Debugging and testing

### Work done so far

- Automated story illustration dataset
  - ImageCLEF 2010 data set (237,434 different images + metadata)
  - 27 stories taken from Aesop's Fable
- Wikipedia 2016 dataset
  - Significantly larger, more heterogeneous (selected from 12+ million pages)
  - Queries = 300 Wikipedia articles
  - Each selected article has between 10 and 30 images

Overview of the automated story illustration task.

Debasis Ganguly, I. Calixto, and Gareth Jones. In Proc. FIRE Workshop, 2015.

### Work done so far

- Survey of existing test collections
- Creation of new test collection based on CiteSeer / CiteSeer<sup>X</sup>
  - CiteSeer<sup>X</sup>: document collection, with cleaner, more reliable metadata
  - Our collection: 600,000 research papers + 2,500 queries
- Standard techniques
  - Collaborative filtering approaches: user × item matrix
  - Content based methods (widely used)
  - Reference Directed Indexing: use "anchor text" to index an article

From a Scholarly Big Dataset to a Test Collection for Bibliographic Citation Recommendation.

Dwaipayan Roy, Kunal Ray, and Mandar Mitra. In AAAI Workshop on Scholarly Big Data, 2016.

An Improved Test Collection and Baselines for Bibliographic Citation Recommendation. Dwaipayan Roy. Proc. ACM CIKM, 2017.