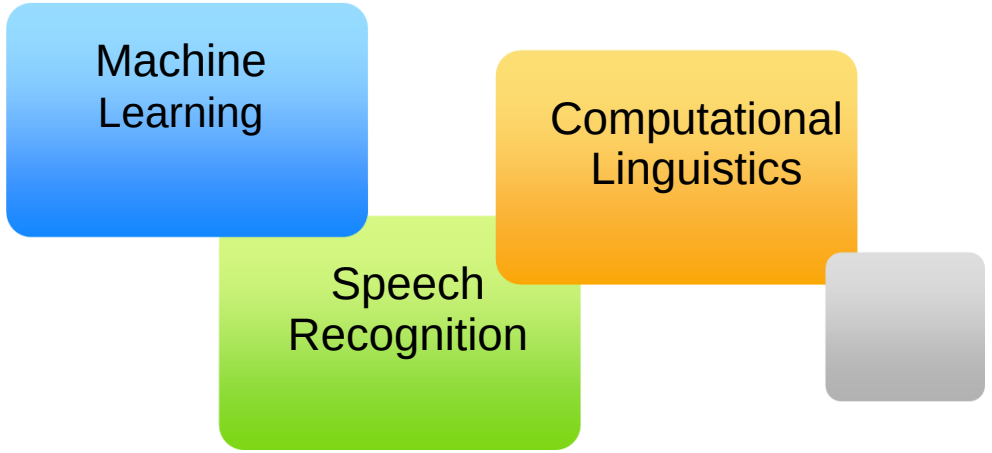
Framework Design for Collaboration in Research

Computer Supported Collaborative Learning
Statistical Machine Learning & its Applications

CMU IIIT-D WINTER SCHOOL 2011



TEAM

Manish Bansal
NIT Rourkela

Shubham Shukla
University of Delhi
Netaji Subhas Institute of Technology

Shaikh Ismail
NIT Rourkela

Objective

- Designing a model that could suggest a possibility of collaboration among two well established researchers having **varied area of research interests.**

- Based on the above model designed, propose a framework that could suggest a **collaborator for guidance in research to any beginner.**

Related Research

- **A Study of Academic Collaboration in Computational Linguistics with Latent Mixtures of Authors (2011)**

Nikhil Johri, Daniel Ramage, Daniel A. McFarland, and Daniel Jurafsky

Topic Modeling and Cosine Similarity

- **Collaborative E-Learning for Remote Education : An Approach For Realizing Pervasive Learning Environments(2006)**

Manikandan, C.; Meenakshi Sundaram, A.S.; Mahesh Babu

K-Means Clustering algorithm

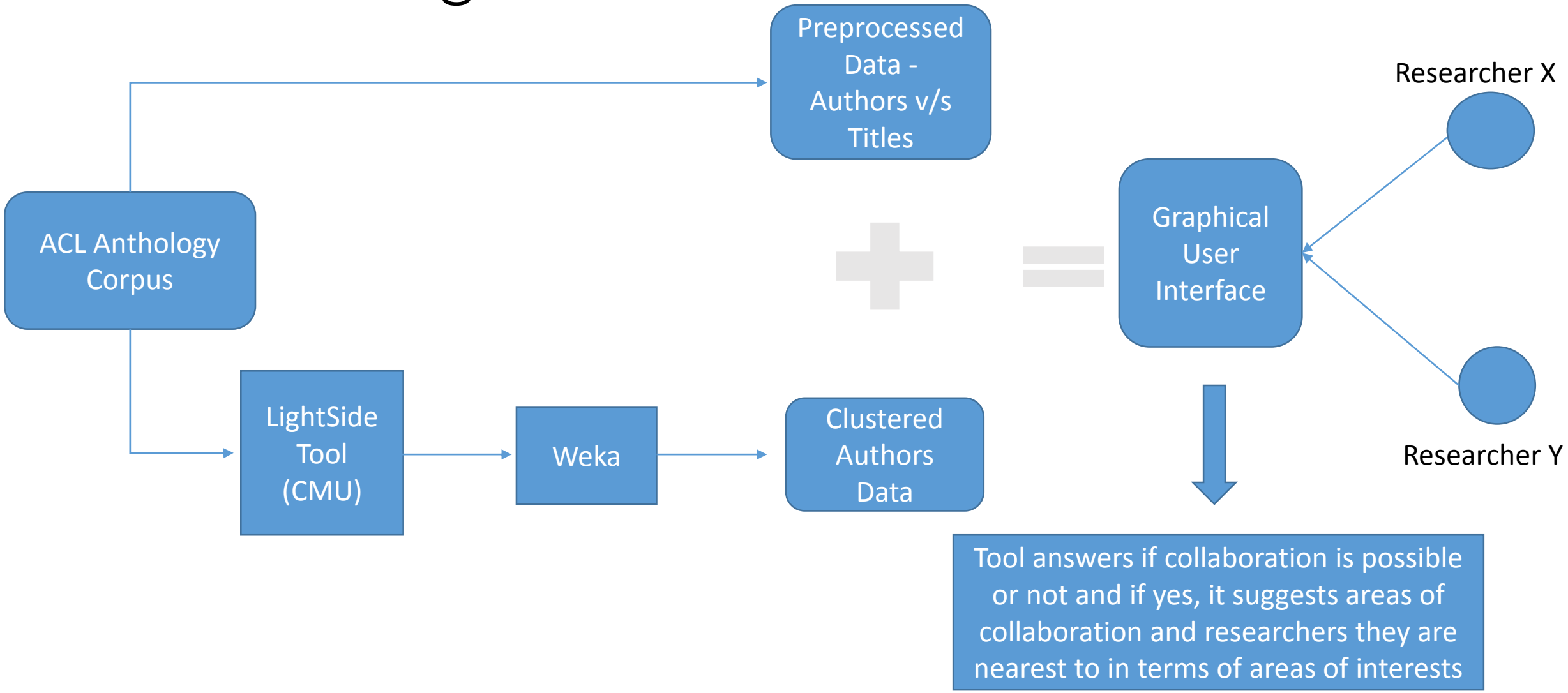
- **Kaleidoscope Concepts and Methods for Exploring the Future of Learning with Digital Technologies (2004)**

Pierre Dillenbourg

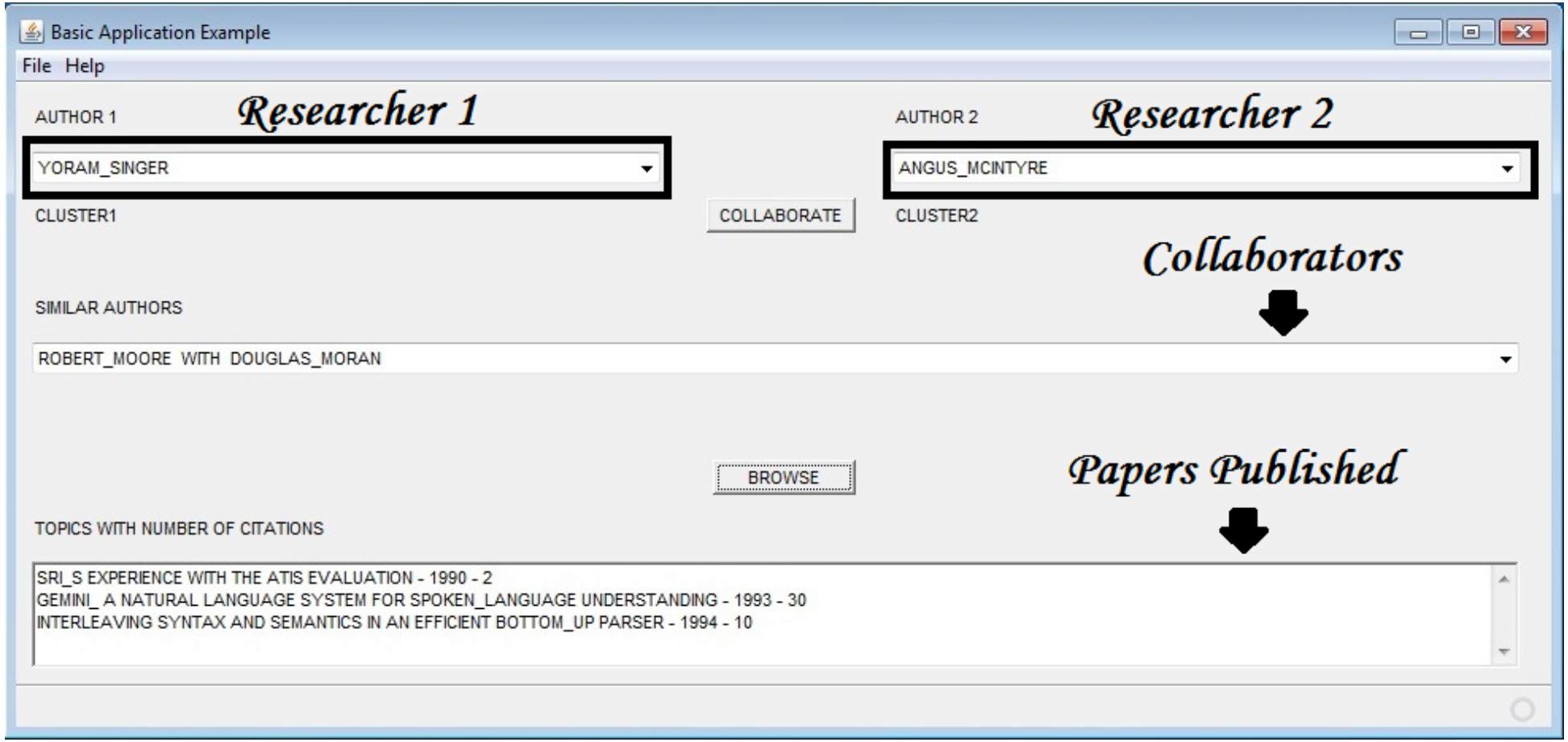
Focus on collaboration among individuals with mixed ability level

WINTER SCHOOL 2011

Model Design

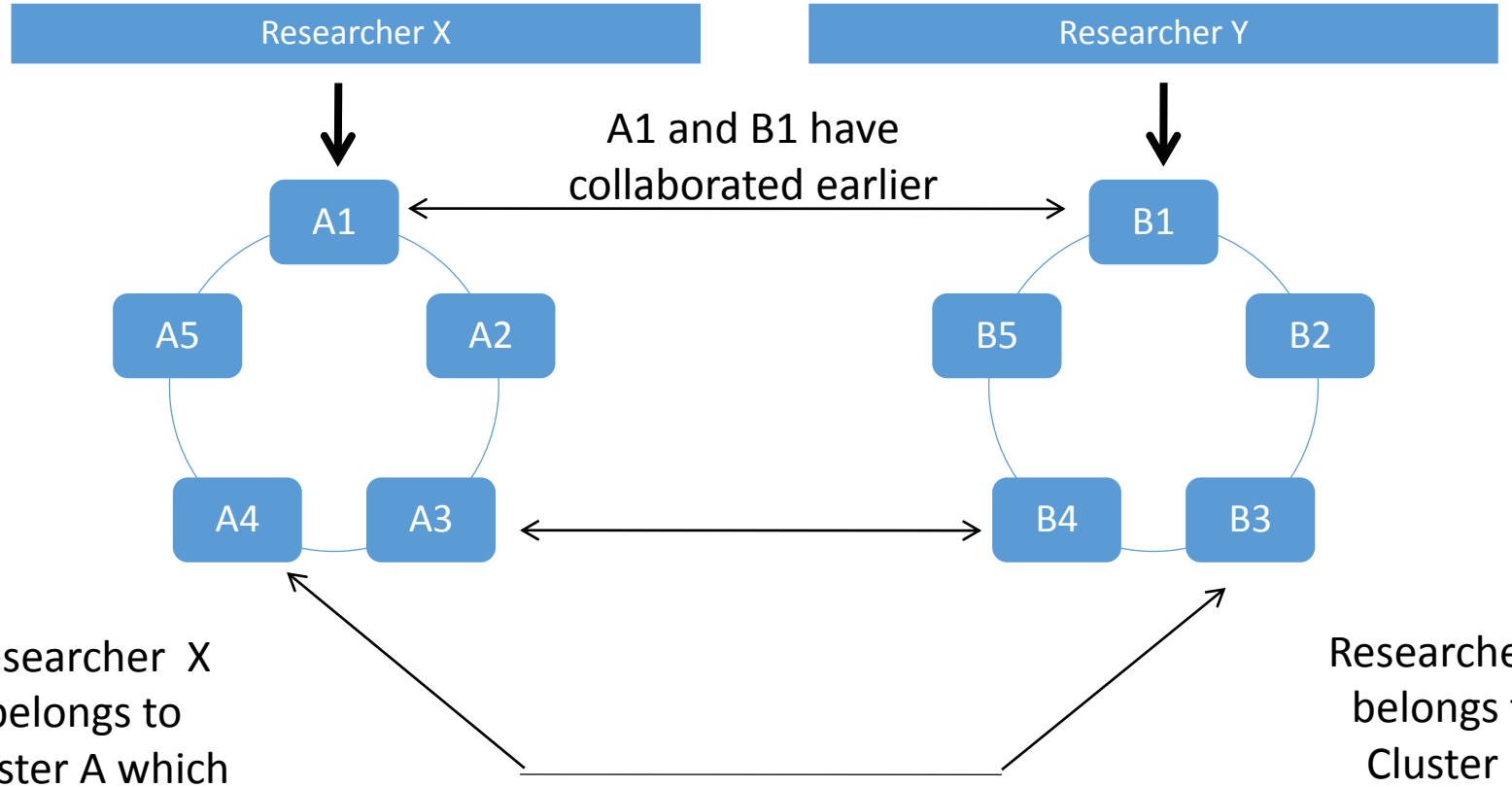


Graphical User Interface Design



WINTER SCHOOL 2011

Key Focus



Researcher X belongs to Cluster A which consists of researchers having same interests

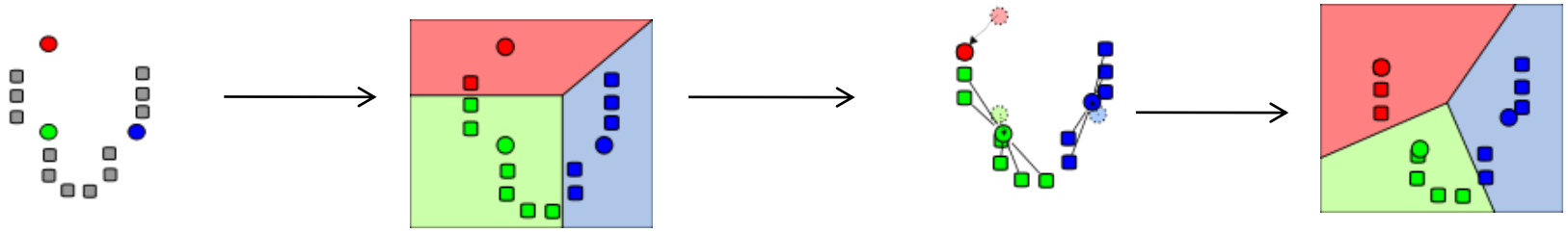
Researcher Y belongs to Cluster B which consists of researchers having same interests.

CLUSTERING !

WINTER SCHOOL 2011

Theoretical Base

- Feature vectors calculated using LightSIDE applying TAGHelper plug-in
- Clustering achieved through K-means clustering algorithm.



For K = 3

∗K means clustering technique forms non-overlapping clusters.

Technical Implementation

❑A good dataset

- ACL Anthology Corpus
- 6000 Research papers
- 5500 Authors
- Data extracted by XML parsing using JAVA , MATLAB and Python.

❑ Issues faced

- Fetching appropriate data: out of 133 XML files, 33 files were corrupt.

Technical Implementation

- Feature extraction : - Input to LightSIDE

Author	Paper Titles (separated by #)
<Author 1>	<Paper 1>#<Paper 2>#< Paper 3>
<Author 2>	<Paper 1>#<Paper 2>#< Paper 3>
…… and so on	<Paper 1>#<Paper 2>#< Paper 3>

- Output from LightSIDE is feature vector of each author (.arff file)

- Implementation of KNN search algorithm to find the nearest researcher based on the input ideas

Technical Implementation

- **Issues faced in extraction and clustering**

  - Author names in XML files were present in erratic format. As a result Arff file generated was not accepted readily by WEKA.
  - To eradicate the same, data collected was cleaned first.
  - Getting feature vectors from LightSIDE was a computationally heavy process even though only paper titles were used.
  - Biased cluster formation

Thank You