



UNIVERSITÀ
DEGLI STUDI
DI MILANO



MEC ITALY

PLAN OF ACTIVITIES

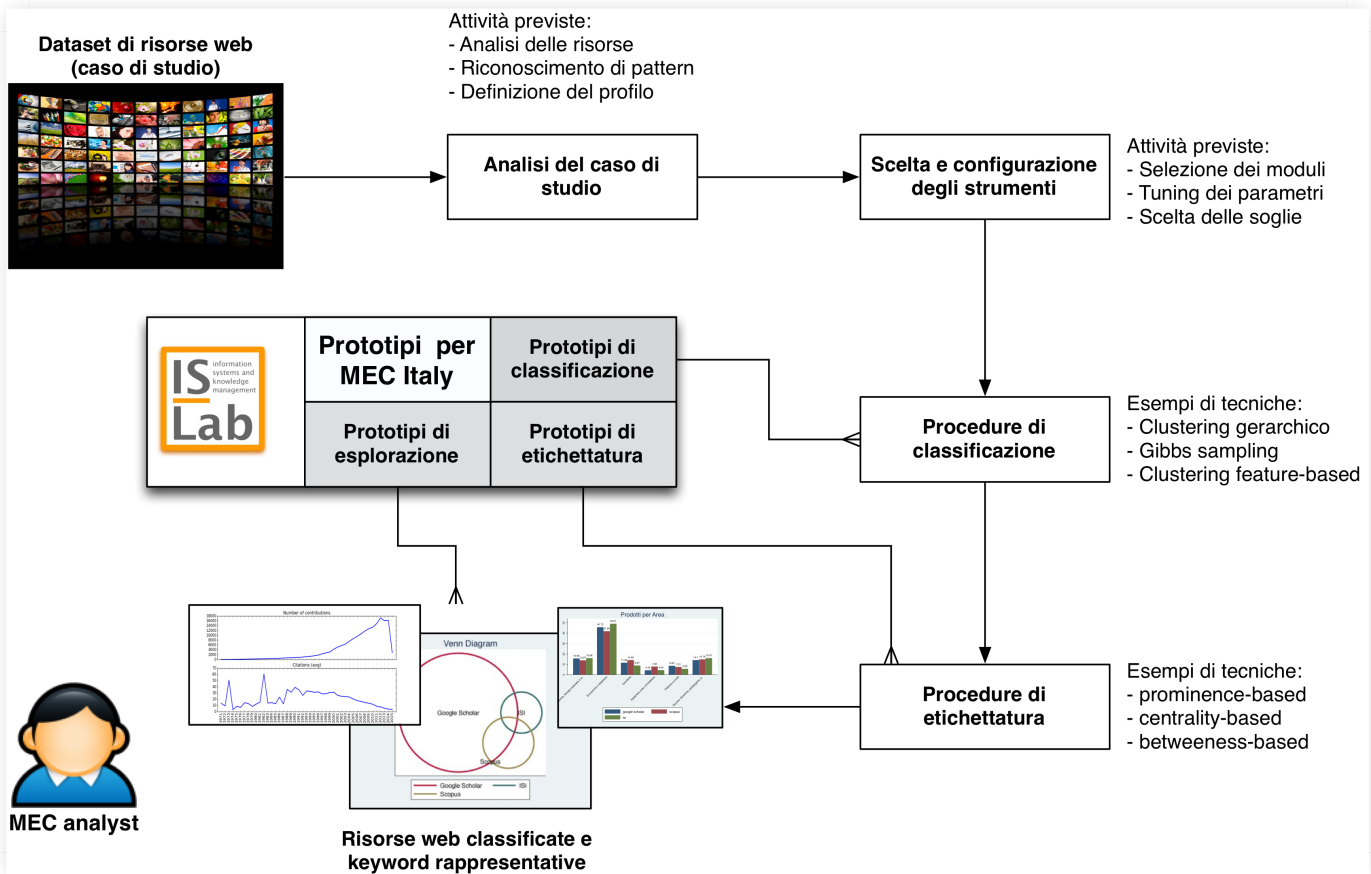
ALFIO FERRARA, STEFANO MONTANELLI

Department of Computer Science

Via Comelico 39, 20135 Milano

{alfio.ferrara,stefano.montanelli}@unimi.it
<http://islab.di.unimi.it/>

REFERENCE ARCHITECTURE



ACTIVITY ORGANIZATION

- **Lecture topics**

- Resource modeling and storage
- Resource indexing
- Text normalization
- Similarity and vector space model
- Resource clustering
- Clustering evaluation
- Topic modeling
- Cluster labeling

- **Lecture modalities**

- Technical issues are presented side-by-side with programming slots for transferring practical skills

TENTATIVE AGENDA

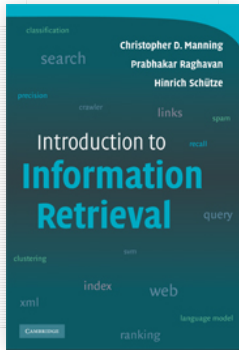
- March 9th
- March 14th
- March 23rd
- March 30th
- April 6th
- April 11th
- April 20th
- April 27th
- ...

TOOLS

The reference programming language will be [Python 2.7](#)

ADDITIONAL TOOLS

- [mongoDB](#)
- [Natural Language ToolKit \(NLTK\)](#)
- [NumPy](#)
- [SciPy](#)
- [gensim](#)
- your curiosity



*Manning, C. D., Raghavan, P., & Schütze, H. (2008).
Introduction to information retrieval (Vol. 1, p. 496).
Cambridge: Cambridge university press. (Available online
at: <http://nlp.stanford.edu/IR-book/>)*

```
@book{manning2008introduction,  
  title={Introduction to information retrieval},  
  author={Manning, Christopher D and Raghavan,  
    Prabhakar and Sch{"u"}tze, Hinrich and others},  
  volume={1},  
  year={2008},  
  publisher={Cambridge university press Cambridge}  
}
```

ANALYSIS OF FIRST ROUND OF EXPERIMENTS

Experiment exploration: [Access to the web prototype](#)

htaccess credentials: username: mec - password: mectag

application credentials: username: mecuser1 - password: mecpsw1