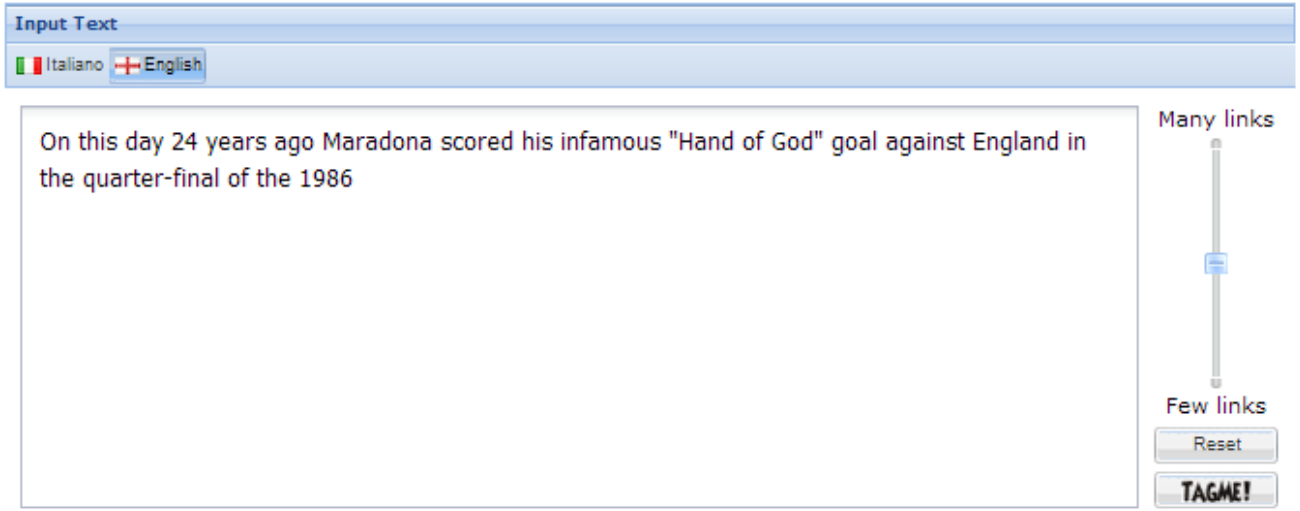
Information Retrieval

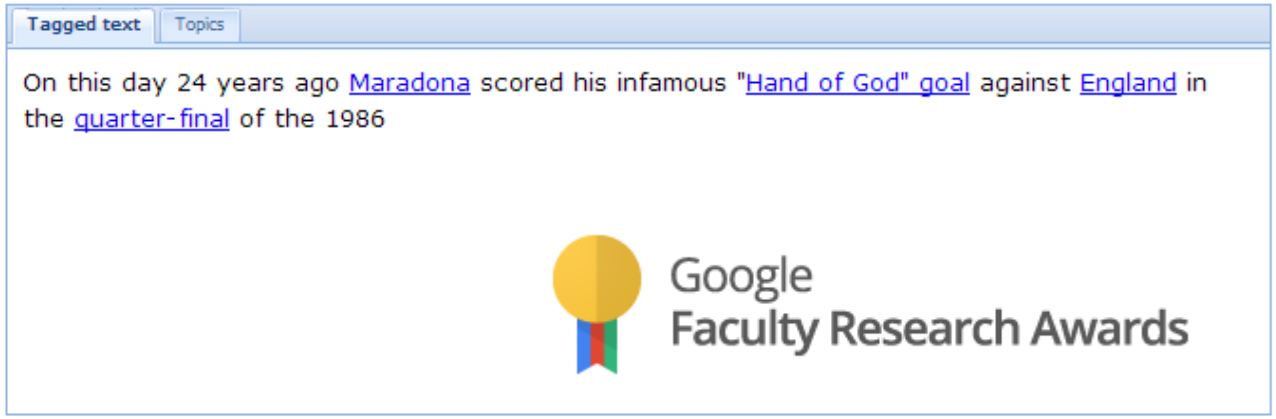
- **Teacher**:
  - Paolo Ferragina
  - paolo.ferragina@unipi.it
- **Semester**: First semester
- **Pre-requisites**: Basics of Algorithms, Algorithm Engineering or Algorithms 2
- **Exam mode**: written *(probably a tiny project on python+Lucene+…)*
- **Area**: Algorithmics

Syllabus

- The main modules of a (Web) Search Engine
  - Crawling
  - Compressed storage
  - Indexing
  - Ranking
  - Query resolution
  - Clustering, classification, recommendation, text annotation, …
- Advanced algorithms and data structures to manage **big data**:
  - (structured and unstructured) Texts
  - (labeled and unlabeled) Graphs
- Probably, an *experience module* on Lucene, TagME, … (in python)

Thesis available TAG·ME





WIKIPEDIA
The Free Encyclopedia