SEMANTIC POINTER

TEXT MINING FOR SEMANTICALLY ENABLED SOCIAL BROWSING

PASCAL Pump Priming
Jure Ferlež
Jožef Stefan Institute, Ljubljana, Slovenia
Presentation Overview

- Project Goal
- Social Browsing Industry
- Idea with Recorded Demo
- Showcased PASCAL Technology
- Dissemination Plan
Social Browsing

- **Main Idea:**
  - One should not be alone while browsing the worldwide web.

- **Existing Solutions**
  - **PEEKKO, GABBBLY, YAPLET**
    - Single web page collaboration context
  - **ME.DIUM**
    - Pattern analysis of the web pages click graph defines the collaboration context
  - **PASCAL - SEMANTIC POINTER**
    - Pattern analysis of the textual content of the web pages defines the collaboration context
Social Browsing Industry
SEMANTIC POINTER

- Web Page Add On
- Allows presence identification and collaboration with visitors to semantically similar content
Welcome to NIPS

NIPS 2008
Tutorials: December 8, 2008
Hyatt Regency, Vancouver, B.C., Canada

Conference Sessions: December 9-11, 2008
Hyatt Regency, Vancouver, B.C., Canada

Workshops: December 12-13, 2008
The Westin Resort & Spa
The Hilton Whistler Resort & Spa

The Foundation: The Neural Information Processing Systems (NIPS) Foundation is a non-profit corporation whose purpose is to foster the exchange of research on neural information processing systems in their biological, technological, mathematical, and theoretical aspects. Neural information processing is a field which benefits from a combined view of biological, physical, mathematical, and computational sciences.

The primary focus of the NIPS Foundation is the presentation of a continuing series of professional meetings known as the Neural Information Processing Systems Conference, held over the years at various locations in the United States and Canada.
Text Mining Technology

- Pattern analysis on textual data for implementing **Semantic Similarity Measure**

- TextGarden technology

- Semantic Similarity Measures on text
  - Very very Simple:
    - BOW, Stop Words, Cosine Similarity
  - More Complicated and Informed using text corpus as Prior Knowledge:
    - TFIDF, LSI (using domain corpus e.g. Videolectures Abstracts)
    - DMOZ, WIKIPEDIA
Putting it all together

SIM = 0.5
SIM = 0.7
SIM = 0.1

TEXT MINING
DISSEMINATION

○ PRECONDITIONS
  ● Finalize the prototype
  ● Perhaps try other GUI approaches

○ PRESENT
  ● Install prototype on various sites:
    • www.videolectures.org
    • www.pascal-network.org?
    • www.ist-world.org
    • …

○ FUTURE
  ● Demonstration to the interested industry
    • Specific companies: Me.dium, Zemanta, …
    • Web Conferences, Helsinki 2008 …
  ● PASCAL2 Harvest project?
CONCLUSION

- Semantic Pointer is a showcase of the PASCAL Text Analysis methods to the Social Browsing Industry

- [http://semanticpointer.ijs.si](http://semanticpointer.ijs.si) (public)
QUESTIONS and IDEAS

- ?
- ?
- ?
- !

hello world! : )
Putting it all together

SIM = 0.85
SIM = 0.66
SIM = 0.44
SIM = 0.34
SIM = 0.23
SIM = 0.81