CS 620 Class Presentation

# Using WordNet to Improve User Modelling in a Web Document Recommender System

**Bernardo Magnini** and **Carlo Strapparava**

Presented by Haifeng He

# Problem

- A recommender system for a Web site of multilingual news
    - **Learns user's interests from the requested pages**
    - **Build a model of the user**
    - **Exploit the model to anticipate which documents in the web site could be interesting for the user**

# Previous Work

- SiteIF, a personal agent for a multilingual news web site
  - **Word–based** (word frequency and co–occurrence)
  - **Not accurate enough**
  - **Misinterpret word sense**

# Main Idea

- **Content-based** document representation
  - □ **Build the user model as a semantic network whose nodes represent sense (not just words)**
  - □ **Retrieve new documents with high semantic relevance with respect to the use model**
  - □ **More accurate and,**
  - □ **independent from the language of the documents browsed(?!).**

- The problems
  - □ **Require a repository for word senses (WordNet)**
  - □ **Word sense disambiguation (WSD)**

# Word Domain Disambiguation

- Sense clustering with domain labels (Magnini and Strapparava, 2000)
  - **Each word has a *domain* label (MEDICINE, SPORT, etc)**
  - **Reduce the WordNet polysemy**
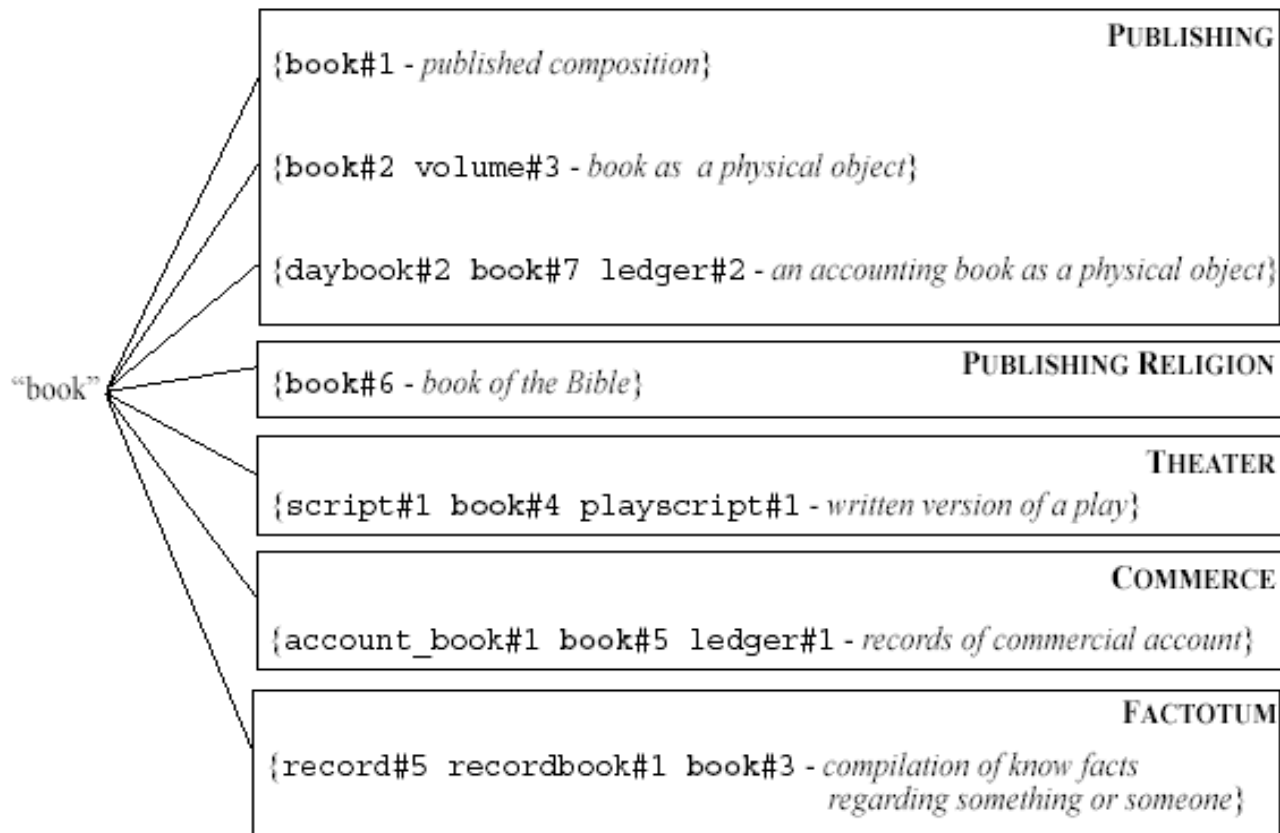  - **Covers only noun synsets now**

# Example



Figure 2: An example of polysemy reduction

# Domain Disambiguation

- Two steps
  - **Given a word, for each domain label of the word, give a score, which is determined by the frequency of the label among the senses**
  - **The domain label with the highest score is selected**
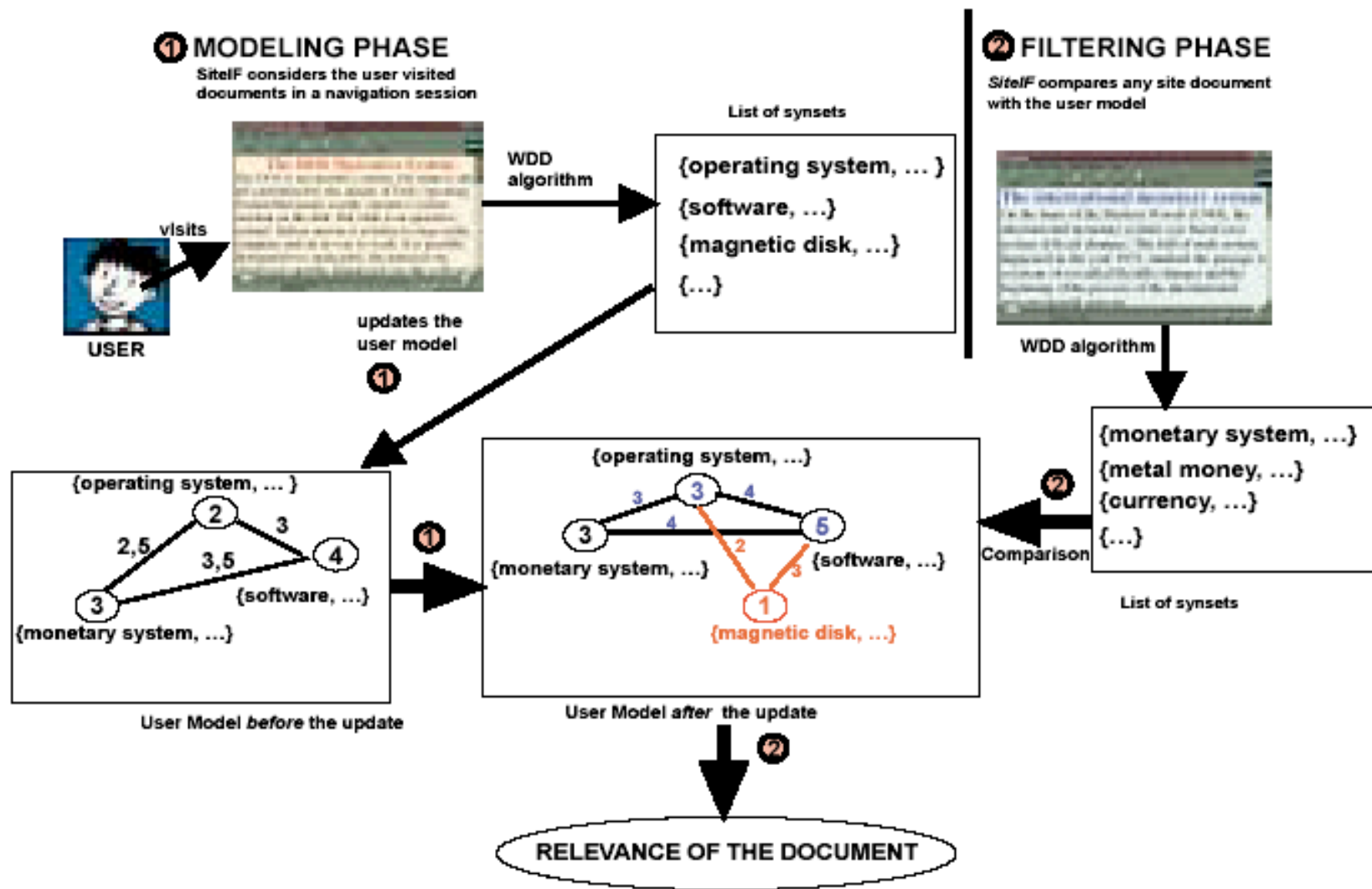  - **.83 accuracy (Magnini and Strapparava, 2000)**

Figure 4: Modelling and Filtering Processes

# Evaluation and Conclusions

- Compare the output of two systems against the judgments of a human advisor

  - □ **Word–based and synset based**

  - □ $H$ **the set of human proposals,** $S$ **the set of the system proposals**

  - □ **Precision =** $\dfrac{|H \cap S|}{S}$ **;    Recall =** $\dfrac{|H \cap S|}{S}$

- Precision increase 34%. Recall increase 15%.