

Ontology Based Approach to Patent Mining for Relating International Patent Classification (IPC) to a Scientific Abstract

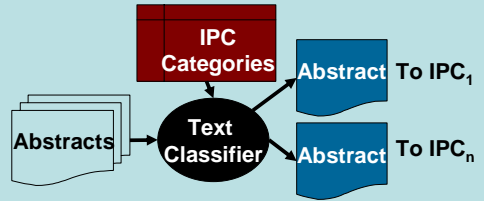
Md. Hanif Seddiqui, Yohei Seki and Masaki Aono
 Toyohashi University of Technology, Aichi, Japan
 hanif@kde.ics.tut.ac.jp, {seki, aono}@ics.tut.ac.jp

Introduction

Motivation:

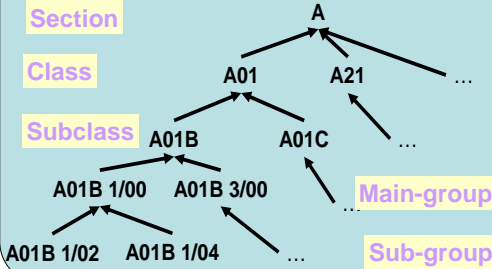
- *Neighboring Semantic (NS)* of
 - Categories (IPCs)
 - Features (Terms)
1. To disambiguate terms, and
 2. To classify text correctly.

Basic Diagram:



Pre-processing

Simple Example of IPC Hierarchy



IPC Ontology

- Create IPC Taxonomy from XML Format
- **Machine Learning Techniques**
 - Remove rare words based on Doc. Freq. (DF)
 - Remove general words based on Cat. Freq. (CF)
 - Create Bag of Words to Each IPC by CHI-Square
- Measure Term-weight by TF-IDF

$$\text{CHI-Square, } \chi^2(t, c) = \sum_{i,j} \frac{O_{i,j} - E_{i,j}^2}{E_{i,j}}$$

Main Processing

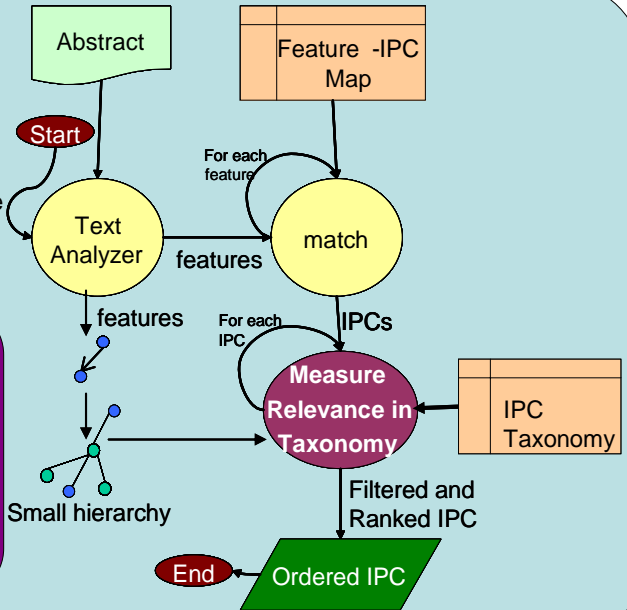
Abstract Text Processing

- Terms/Features Exaction
 - ✓ Text Analyzer
- Inter-feature Relation Extraction
 - ✓ Using Ontology Learning Technique

Probable IPC:

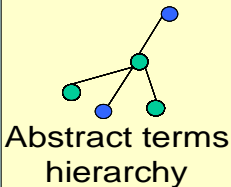
$$\Phi_c = \sum_{f \in a} w(f, c) * TF(f, a)$$

Feature f , belongs to abstract, a .
 $w(f, c)$ is the term/feature weight
 $TF(f, a)$ is the term frequency

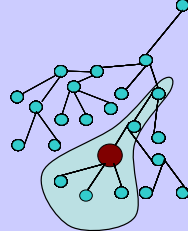


Use of Ontology

Feature NS



Category NS: Bag of Word Taxonomy



Measure

Relevance in Taxonomy

- Let Probable IPC as seed-point
- Look-up abstract features around seed-point of Category NS
- Check availability of semantic relation using aligning technique
- Extract the more specific IPC

Discussion

- ◆ Our Result at NTCIR-7 was poor, because-
 - ☒ We used IPC ontology version 8, however, the experiment data were based on IPC version 6
 - ☒ We did not remove rare-words
 - ☒ We did not measure the closeness of words with the IPC (Bag of Word to IPC)
- ◆ We started modifying our system

Conclusion

- ◆ Use of Semantic Technology in Text Classification
- ◆ Use of Our Ontology Alignment Technique
- ◆ Two Level of Neighboring Semantic for WSD