Vienna, 8/9 November 2007

Patent Analysis

- The PATExpert View -

Information Retrieval Facility Symposium

Leo Wanner

patexpert@upf.edu

http://www.patexpert.org

Tools for Management and Research in the Patent Domain

A patent processing work bench must be able to

- Rephrase a patent in terms the reader can understand
- Provide the essence of a given patent (and pin down its novelty)
- Determine the (essential) difference between given patents
- Determine the general trend with respect to the change of inventions in a given field
- Allow for content-oriented search
- Assess the value of a patent

What kind of analysis do we need?

… basically, the whole range of linguistic analyses:

- Word level analysis
  For all tasks
- Text structure analysis
  For (multilingual) gist synthesis, search, …
- Discourse structure / anaphoric structure analysis
  For gist synthesis, paraphrasing, search, …
- Syntactic structure analysis
  For paraphrasing, gist synthesis, content distillery, …
- Semantic analysis
  For content distillery, search, gist synthesis, …

Simplification: A Necessary Preprocessing Stage

Before many analysis tasks are carried out, a drastic simplification of the linguistic style of the material is desirable

- Cut each single sentence into a number of separate sentences taking into account surface-oriented criteria
- Transform FOR – gerund constructions into finite clauses
- Eliminate “excessive” anaphora markers (such as SAID ...), substututing it where necessary a definite ART

Text Structure Analysis

Overall patent document structure:

- ⇨ title, abstract, claims, description, ...

Patent claim structure:

- ⇨ independent / dependent claims
- ⇨ interclaim references
- ⇨ paragraph / sentences

- Identify the text structure using titles, cue words, numbering, etc., i..e, surface-oriented criteria

Example

Discourse and Reference Structure Analysis

Discourse Structure: Discourse tree in the sense of the *Rhetorical Structure Theory*

Reference Structure: Anaphoric lexical chains

- Cue words serve as the main means to detect discourse relations
  - ⇨ "so that" → MANNER
  - ⇨ "wherein" → ELABORATION
  - ⇨ "for X_{gerund}" → PURPOSE
  - ⇨ "comprising" → $ELABORATION_{part\text{-}whole}$
  - ⇨ ...
- Anaphoric references are, as a rule, realized as repetitions (sometimes with the SAID modifier) and thus "easy" to detect

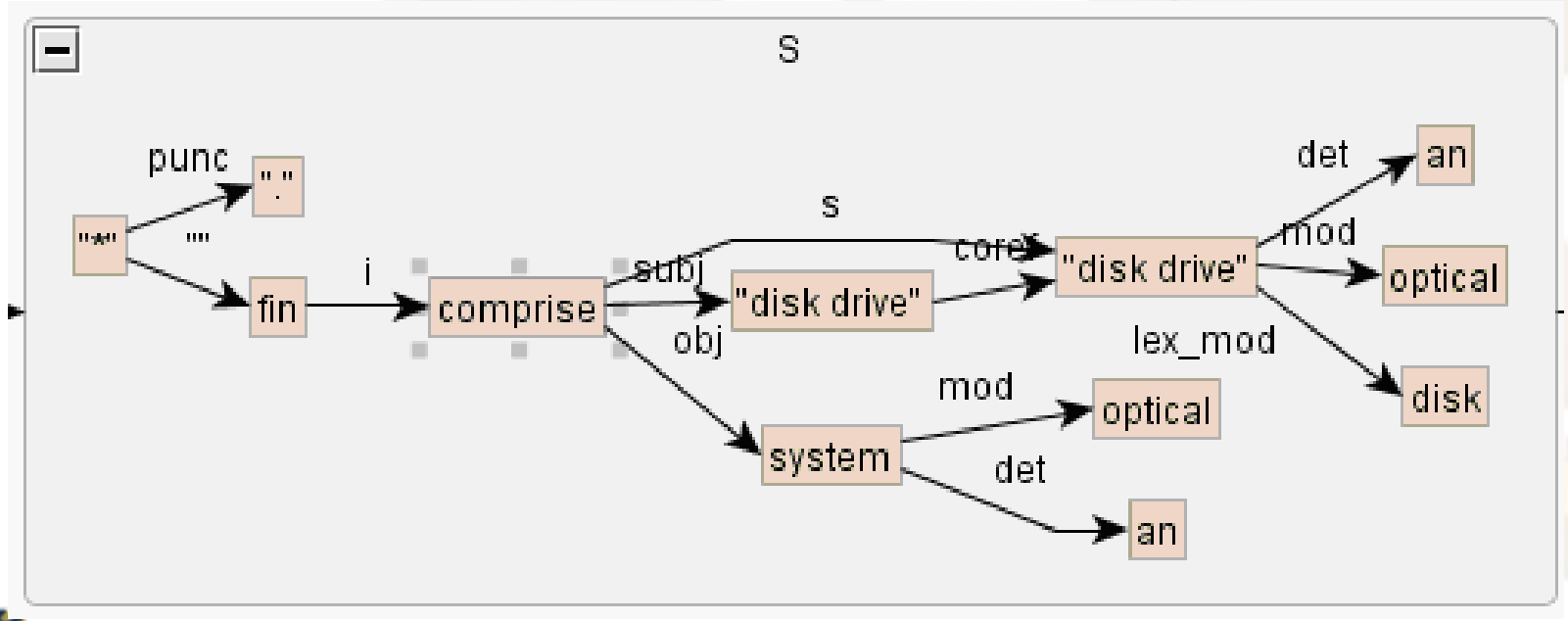
Example

Multiword Unit Identification

- Create a stop unit list
  - ⇨ All one-character tokens
  - ⇨ Domain-specific functional words
- Get a list of most frequent *n*-grams in the patent corpus
  - ⇨ Build list of all possible *n*-grams ranked by frequency
- Filter out irrelevant *n*-grams
  - ⇨ *n*-grams containing stop units (e.g., DET N)
  - ⇨ *n*-grams containing certain PoS-patterns (e.g., V_{fin} N)
- Obtain multiword units
  - ⇨ any *n*-gram which occurs at least N_1 times in the corpus and/or at least N_2 times in a patent; e.g., “spindle motor”, “light beam”, “optical pickup”, …

Syntactic Structure Analysis

- Dependency-based parsing (of previously simplified material) using MiniPar
- Mapping to other dependency formats (when required)



Key Concept Analysis

Concept:

any “base form” mono- or multiword term

- A “base form” mono-/multiword term captures all morphological and lexical variances of equivalent terms (e.g., *CD*, *compact disk*, *compact disc*)

Key Concept:

any concept of sufficient relevance

- A concept is of relevance if it appears among the *N* first concepts ranked with respect to the tf*idf weight metric

<optical disc apparatus>, <recording>, <recording medium>, <information face>, <light beam>, <focus control system>, ...

Semantic Analysis

Frame-based analysis (one of the possible strategies)

- ⇨ A frame is a conceptual structure encoding an assertion and involving a number of participants (aka *frame elements*, *FEs*)
- ⇨ Each frame is expressed by specific lexical units; an LU expressing a FE is related to its frame name by a specific grammatical function

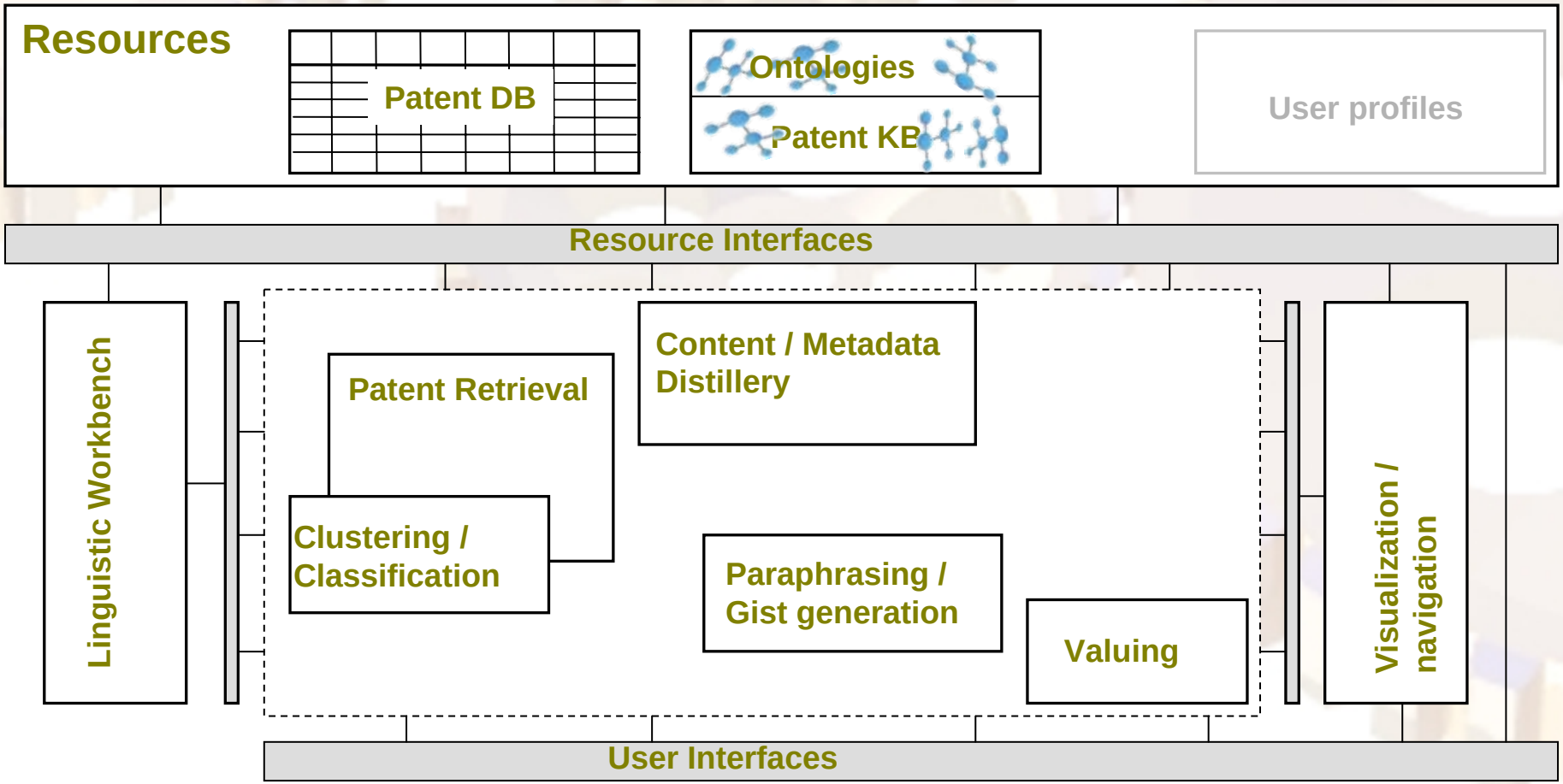
- Parse a patent with MiniPar
- Recognize and extract frames and their elements

Frame: INCLUSION	LU *comprise*:	subj →whole; obj →part
FE 1: whole	LU *include*:	subj →whole; obj →part
FE 2: part	LU *be_equipped*:	subj →whole; mod (*with*) →part

Vienna, 8/9 November 2007

The Use of Analysis Techniques in PATExpert, some examples

The PATExpert-Service



Content Distillery in PATExpert

- Shallow content distillery
  - ⇨ Based on key-concept analysis / extraction
  - ⇨ Exploits basic ling. analysis (PoS-tagging, lemmatization, multiword detection)

US6011762 <light beam>, <recording medium>, <information face>, <focus jumping>, <focus control>, …

- Deep content distillery
  - ⇨ Based on frame analysis
  - ⇨ Exploits complex ling. analysis (parsing, word sense disambiguation)

*Fig 1. is a **view** **showing** an example of an **optical recording and reproducing apparatus***

I_10_5-10.5	rdf:type	pulo:TechnicalDrawing
I_10_16-10_16	rdf:type	ordo:apparatus
I_10_5-10-5	pulo:InfoSubj	I_10_16-10_16

Patent (Claim) Paraphrasing and Summarization in PATExpert

- Paraphrasing and shallow summarization
  - ⇨ Based on text structure and discourse structure analyses
  - ⇨ Based on syntactic analysis
  - ⇨ Exploits sentence simplification
  - ⇨ Exploits anaphoric reference resolution (intra- and inter-claim)
- Deep summarization
  - ⇨ Based on semantic analysis
  - ⇨ Exploits text structure and discourse structure analyses

Optical disk reading apparatus

EP0548937A1

Overview | Claims | Description | Inventors | Applicants | Classes | Classification | Events | Family | Priorities | Pictures | Key Conce

Paraphrase Claim Level: 0% Discourse Level: 0% Summarization

Paraphrased claims

No.	Text
1	The optical disk drive comprises a laser light source, a signal processing circuit, an optical system, and a detection means. The laser light source emits a laser beam. The optical system converses the laser beam from the laser light source on a signal plane of optical disk on which signal marks are formed. Furthermore, the optical system transmits the light reflected from the signal plane. The optical disk drive also comprises one or more optical components. These components are arranged in the optical path between the laser light source and the optical disk, for making the distribution of the laser beam converged by the conversing means located on a ring belt just after the passage of an aperture plane of the optical system. This is so that, the detection means detects the light reflected from the optical disk, and the signal processing circuit generates a secondary differential signal. For this, it differentiates the signals detected by the detection means. It also detects the edge positions of the signal marks. In order to do so, it compares the secondary differential signal with a detection level.

	No.	Text
☑	1	An optical disk drive comprising : a laser light source for emitting a laser beam ; an optical system for conversing the laser beam from the laser light source on a signal plane of optical disk on which signal marks are formed and for transmitting the light reflected from the signal plane ; one or more optical components , arranged in the optical path between the laser light source and the optical disk , for making the distribution of the laser beam converged by the conversing means located on a ring belt just after the passage of an aperture plane of the optical system ; a detection means for detecting the light reflected from the optical disk ; and a signal processing circuit for generating a secondary differential signal by differentiating the signals detected by the detection means and for detecting the edge positions of the signal marks by comparing the secondary differential signal with a detection level .
		The optical disk drive according to Claim 1 , wherein said signal processing circuit comprising : a differentiating means for generating the primary differential and the secondary differential of the readout

Optical disk reading apparatus

EP0548937A1

Overview | Claims | Description | Inventors | Applicants | Classes | Classification | Events | Family | Priorities | Pictures | Key Concepts

Paraphrase Claim Level: 25% Discourse Level: 25% Summarization

Summarized claims:

An optical disk drive comprises a laser light source, an optical system, one or more optical components, a detection mea
and a signal processing circuit.

No.	Text
1	An optical disk drive comprising : a laser light source for emitting a laser beam ; an optical system for conversing the laser beam from the laser light source on a signal plane of optical disk on which signal marks are formed and for transmitting the light reflected from the signal plane ; one or more optical components , arranged in the optical path between the laser light source and the optical disk , for making the distribution of the laser beam converged by the conversing means located on a ring belt just after the passage of an aperture plane of the optical system ; a detection means for detecting the light reflected from the optical disk ; and a signal processing circuit for generating a secondary differential signal by differentiating the signals detected by the detection means and for detecting the edge positions of the signal marks by comparing the secondary differential signal with a detection level .
2	The optical disk drive according to Claim 1 , wherein said signal processing circuit comprising : a differentiating means for generating the primary differential and the secondary differential of the readout signal detected by said detection means ; a first gate which opens when the primary differential signal exceeds a higher gate level at the plus side ; a second gate which opens when the primary differential signal exceeds a lower gate level at the minus side ; and a cross detector for detecting the cross point of the secondary differential signal with the detection level when the first or second gate is open .
3	The optical disk drive according to Claim 1 , wherein said one or more optical components is a shading member arranged near the optical axis around the aperture plane of said optical system .
4	The optical disk drive according to Claim 1 , wherein said one or more optical components is a transparent conical body arranged near the optical axis around the aperture plane of said optical system .
	An optical disk drive comprising : a laser light source for emitting a laser beam ; an optical system for

Patent Retrieval in PATExpert

- Classic (keyword-based search)
  - ⇨ Exploits the results of multiword unit identification
  - ⇨ Exploits the results of ling. analysis (by indexing patent claim summaries instead of full patents)
- Semantic search
  - ⇨ Exploits the results of semantic analysis (by allowing for content relations and concepts in search queries)
- Document similarity search
  - ⇨ Exploits the results of multiword unit identification
  - ⇨ Exploits the results of semantic analysis

The experience teaches us:

- If we want to offer intelligent patent processing techniques, robust linguistic analysis at all strata of the linguistic model is essential

- Analysis of patent material may well pose different challenges that analysis of general discourse