LEXICAL SEMANTICS AND EXPERT LEGAL KNOWLEDGE TOWARDS THE IDENTIFICATION OF LEGAL CASE FACTORS

JURIX 2010

DEC. 17, 2010, LIVERPOOL

Adam Wyner - University of Liverpool
Wim Peters - University of Sheffield
OVERVIEW

- Case-based reasoning.
- Goals.
- Decomposition of factors.
- Outline the tool *how-to* and give some sample output.
Argue a current undecided case with respect to precedents.

Cases analysed in terms of factors.

Compare and contrast the cases in terms of the factors; where the same factors hold, the same decision should be applied; where other factors hold, reason by a counterbalancing of factors.

Identifying factors on basis of facts is a difficult, skilled, time-consuming, error-prone manual task. Knowledge bottleneck. The form-meaning problem.
GOALS

- Provide a semi-automated legal text analysis tool which incorporates lexical semantics and expert legal knowledge to support the legal professional in the identification of legal case factors.
  - For users: flexible, interactive, constructive, and collaborative.
  - For developers: open, modular, adaptive, and verifiable; gold standard development.
  - Explicit justification and verification of factor in text.
  - "Crowd-sourcing" annotation and development.
GOALS

Not:

- Fully automatic annotation (Wyner and Peters 2010).
- Machine learning (Ashley and Brunninghaus 2009).
- Case-based reasoning, but the basic input.
- End of lawyers.
Based on the CATO corpus of +140 cases about trade secrets. Not all accessible, so working with a subset of +30 cases. Issues of building a case base.

Our (available upon request) case base links to the source text and associates them with factors (derived from the literature).
Aleven (1997) has 27 base level factors. Factors are presented descriptively in text (our adaptation).

- F1 Plaintiff-disclosed-information-in-negotiations
  - Favours defendant.
  - Plaintiff disclosed information during negotiations with defendant. The defendant fairly obtained the information and the plaintiff was not interested to maintain the information as a secret.
  - Applies if the plaintiff disclosed the information to defendant during negotiations for a joint venture, licensing agreement, sale of a business, etc.
  - Does not apply if the defendant learned the information while employed by plaintiff.
Our approach is *bottom up*, conceptually organised, using linguistic and computational linguistic techniques.

- List elements or words of similar meaning.
- Associate each list with a 'cover concept'.
- Annotate the corpus with the concepts.
- Search the corpus for concepts or combinations of concepts.
- Results are patterns of occurrences of concepts.
For the development of the tool in the General Architecture for Text Engineering (GATE), we developed factoroids.

A factoroid is a conceptual cover. They are not the factors themselves, but are the linguistic terms and relations which comprise the base level factors.

Decomposition.

From the factor presentations (e.g. of F1), manually extract and lemmatise the most salient terms, e.g.

- plaintiff, disclose, product, information, negotiation, obtain, fair means
Consult WordNet (online Thesaurus) and manually identify relevant, synonymous words. For disclose, we find:

- announce, betray, break, bring out, communicate, confide, give away, impart, inform

GATE uses such lists as a bottom level of the annotation process.
We are looking for facts (events, states, actions and related to entities in grammatical roles) which are indicative of factors.

Further linguistic information to support factor identification.

VerbNet gives the range of required syntactic arguments and the semantic properties of verbs.

Imported VerbNet into GATE as lists to be used in annotating verbs.
Syntactic - subject, verb, object, indirect object, noun phrase, verb phrase, etc. *The plaintiff received the information from the defendant*, where the verb requires a subject noun phrase.

Thematic roles - Agent, Theme, Recipient, and others (30), which are the semantic properties that arguments in the sentence bear. *The plaintiff received the information from the defendant*, where *the plaintiff* is the Agent (doer of action) and *the defendant* is the Recipient.
Selection restrictions - abstract, body_part, animate, organisation, (36). These restrict arguments bearing thematic roles to certain sorts. *The rainstorm received the information from the defendant* is odd since the subject is not an animate agent.

Semantic predicates - command, confront, forbid, group, (144). Abstract semantic properties of classes of verbs and for inference. *The plaintiff received the information from the defendant*, where *receive* includes *has_possession*. 
GATE is a framework for language engineering applications, which supports efficient and robust text processing.

The tools are formed into a pipeline (a sequence of processes):
- sentence splitters, tokenisers, part-of-speak tagger, morphological analyser, gazetteer lists, and Java Annotation Patterns Engine (JAPE) rules.

JAPE rules use basic or previous annotation and context to produce annotations.
JAPE RULES

Rule: DisclosureFactor01
Priority: 5

{Lookup.majorType == "disclosure"}::temp
-->::temp Disclosure = {rule = "DisclosureFactor01"}

Rule: DisclosureInformationTempXY
Priority: 5

{Disclosure}
({Token, !Split})*
{Information}::temp
-->::temp DisclosureInformationTempXY = {rule = "DisclosureInformationTempXY"}
After annotating, can query for strings and annotations (using ANNIC, based on Lucene):
- DisclosureFactoroid will tag all those tokens in a corpora which bear that annotation, e.g. confide, impart....

We have +40 gazetteer lists along with related JAPE rules for legal domain knowledge and factors:
- UseHave, Confidential, Disclosure, Disseminate, FormEmployee, Hire, Information, Know, Negotiate, Outsider, SecureInformation, LegalParties...
SAMPLE - LINEAR AND DISTRIBUTED

Context: also strongly urges that the defendant failed to disclose information which it should have disclosed

Token:

InformationFactoroid:

LegalPartiesFactoroid:

NeglectSemanticPred:

DiscloseFactoroid:

Left context | Match | Right context | Features
--- | --- | --- | ---
also strongly urges that the | defendant failed to disclose information | which it should have disclosed | Lookup={majorType=de...length=11, |
also strongly urges that the | defendant failed to ...tion which it should have disclosed as an agent | | lookup={majorType=de...length=6, |
also strongly urges that the | defendant failed to ...information which it | should have disclosed as an | lookup={majorType=de...length=2, |
also strongly urges that the | defendant failed to ...se information which | it should have disclosed as | lookup={majorType=de...length=5, |
SAMPLE (LINEAR)
Start with a simple search, get results, inspect the context for suggestions for possible further refinements, add additional terms, etc.

- \{LegalPartiesFactoroid\}
- \{LegalPartiesFactoroid\} (\{Token\})*3 \{NeglectSemanticPred\}
- \{LegalPartiesFactoroid\} (\{Token\})*3 \{NeglectSemanticPred\} (\{Token\})*5 \{DisclosureFactoroid\} (\{Token\})*3 \{InformationFactoroid\}
defendant obtained trade secrets from an engineer formerly employed by plaintiff under a contract not to reveal such secrets.
In the instant case the Christophers deliberately flew over the DuPont plant to get pictures of a process which DuPont had attempted to keep secret. The Christophers delivered their pictures to a third party who was certainly aware of the means by which they had been acquired and who may be planning to use the information contained therein to manufacture methanol by the DuPont process. The third party has a right to use this process only if he obtains this knowledge through his own research efforts, but thus far all information indicates that the third party has gained this knowledge solely by taking it from DuPont at a time when DuPont was making reasonable efforts to preserve its secrecy. In such a situation DuPont has a valid cause of action to prohibit the Christophers from improperly discovering its trade secret and to prohibit the undisclosed third party from using the improperly obtained information.

We note that this view is in perfect accord with the position taken by the authors of the Restatement. In commenting on improper means of discovery the savants of the Restatement said:

"If improper means of discovery. The discovery of another's trade secret by improper means subjects the actor to liability independently of the harm to the interest in the secret. Thus, if one uses physical force to take a secret formula from another's pocket, or breaks into another's office to steal the formula, his conduct is wrongful and subjects him to liability apart from the rule stated in this Section. Such conduct is also an improper means of procuring the secret under this rule. But means may be improper under this rule even though they do not cause any other harm than that to the interest in the trade secret. Examples of such means are fraudulent misrepresentations to induce disclosure, tapping of telephone wires, eavesdropping or other espionage. A complete catalogue of improper means is not possible. In general they are means which fall below the generally accepted standards of commercial morality and reasonable conduct." Restatement of Torts § 757, comment f at 10 (1939).
FUTURE WORK

- Annotation experiments with users to study the utility of the search terms to help users identify factors.
- Problems of word sense selection.
- Mapping of lexical semantics to syntactic arguments.
- Integrate syntactic parsing, ontologies, and existing linguistic and terminological knowledge bases.
THANKS!