

WEB-BASED SOFTWARE TOOLS TO SUPPORT STUDENTS' EMPIRICAL STUDY OF THE LAW

BILETA 2011

APRIL 11, 2011, MANCHESTER

Adam Wyner - University of Liverpool

Wim Peters - University of Sheffield

Fiona Beveridge - University of Liverpool

OVERVIEW

- Empirically studying cases.
- Case-based reasoning.
- Goals.
- Decomposition of factors.
- Outline the tool *how-to* and give some sample output.
- Prospects for legal education.

EMPIRICAL CASE STUDY

- Applying text analytic/corpus linguistic techniques to legal cases for content analysis.
- Advantages:
 - Identifies elements in a legal case.
 - Supports search for these elements.
 - Collaborative and online work.
 - Jurimetrics - correlations between jurisdictions, justices, representation, facts, etc.
- ***How to*** use automated or distributed text analytics to facilitate these tasks?

LEARNING TO READ CASES

- Simple elements:

parties, jurisdiction, legislative and case citations, cause of action, ratio decideni....

- Complex elements such as material facts, argument patterns, rules:

"A stranger to a trust should only be held to account as a constructive trustee if he or she acted dishonestly."

- 1. Where in the cases is the term 'stranger to a trust' defined?
- 2. What material facts in the cases are used to indicate 'acting dishonestly'?
- 3. What legislation or laws are cited which bear on the rule?

WHAT IS NEEDED?

- Some details can be searched for using Google-like searches for terms, but text analysis is a skill acquired over time.
- What support tools can we provide to students to help them hone in on relevant material and develop their analytic capabilities?
- In the following, we describe development of a software tool to help students identify relevant elements from a case. Done on the web, this leads to collaborative analysis of cases.
- Look at material facts (aka factors).

CASE-BASED REASONING

- 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 Bruninghaus 2009).
- Case-based reasoning, but the basic input.
- End of lawyers.

CORPORA

- 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).

FACTORS

- 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.

OVERVIEW OF APPROACH

- 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.

ANALYSIS OF FACTORS INTO FACTOROIDS

- 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

ANALYSIS OF FACTORS INTO FACTOROIDS

- 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.

VERBAL SYNTAX AND SEMANTICS

- 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.

VERBAL SYNTAX AND SEMANTICS

- **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.

VERBAL SYNTAX AND SEMANTICS

- **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*.

GENERAL ARCHITECTURE FOR TEXT ENGINEERING (GATE)

- 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"}
```

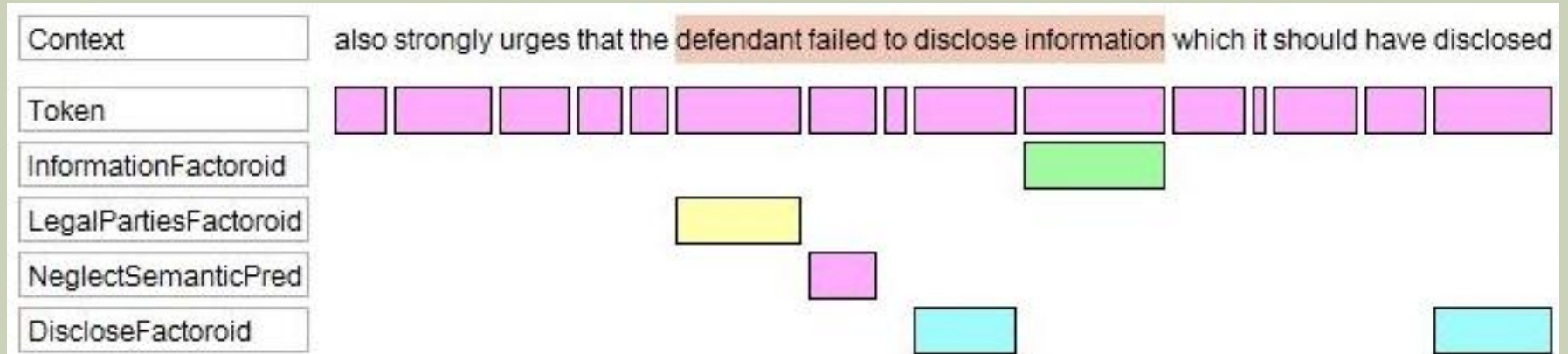
```
Rule: DisclosureInformationTempXY
Priority: 5
{
{Disclosure}
{{Token, !Split}}*
{Information}
}:temp
-->
:temp.DisclosureInformationTempXY = {rule = "DisclosureInformationTempXY"}
```

ANNOTATIONS

- 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...

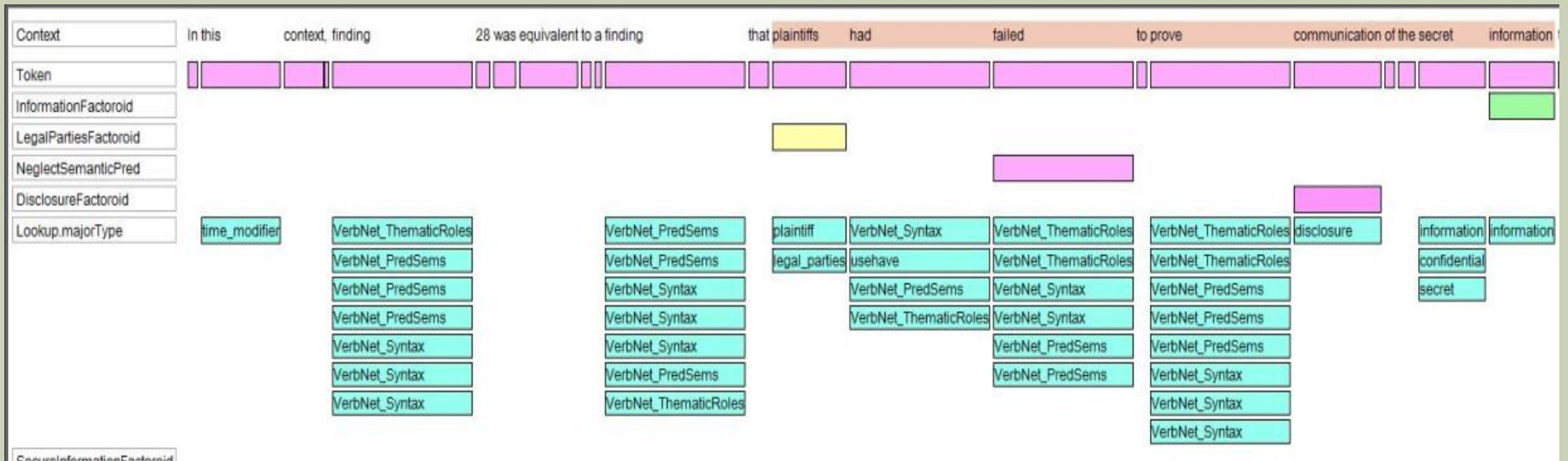
SAMPLE - LINEAR AND DISTRIBUTED

{LegalPartiesFactoroid} ({Token})*3 {NeglectSemanticPred}({Token})*5{DiscloseFactoroid}({Token})*3{InformationFactoroid}



Left context	Match	Right context	Features
also strongly urges that the	defendant failed to disclose information	which it should have disclosed	Lookup={majorType=de...ength=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)



INCREMENTAL, INTERACTIVE, CONSTRUCTIVE

- 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}}

FURTHER EXAMPLE

{ConfidentialFactoroid}{Token}*10{HireFactoroid}{Token}*5{LegalPartiesFactoroid}{Token}*10{DiscloseFactoroid}

Corpus: Entire datastore Annotation set: All sets

Results: Context size:

Search Clear Next page of 50 results

Context defendant obtained trade secrets from an engineer formerly employed by plaintiff under a contract not to re..eal such secrets.

RecipientThematicRole

ConfidentialFactoroid

HireFactoroid

LegalPartiesFactoroid

DiscloseFactoroid

The screenshot shows a search interface. At the top, there is a search bar containing the query: {ConfidentialFactoroid}{Token}*10{HireFactoroid}{Token}*5{LegalPartiesFactoroid}{Token}*10{DiscloseFactoroid}. To the right of the search bar are controls for 'Corpus' (set to 'Entire datastore') and 'Annotation set' (set to 'All sets'). Below these are 'Results' and 'Context size' indicators, and buttons for 'Search', 'Clear', and 'Next page of 50 results'. The main area displays a search result snippet: 'Context defendant obtained trade secrets from an engineer formerly employed by plaintiff under a contract not to re..eal such secrets.' Below the snippet is a list of factoroid types: 'RecipientThematicRole', 'ConfidentialFactoroid', 'HireFactoroid', 'LegalPartiesFactoroid', and 'DiscloseFactoroid'. Colored boxes are placed on the snippet to indicate matches: a blue box under 'defendant', a green box under 'trade secrets', a cyan box under 'from an engineer', a blue box under 'formerly employed', a cyan box under 'by plaintiff', a green box under 'under a contract', a cyan box under 'not to re..eal', and a blue box under 'such secrets.'

NON-LINEAR VIEW

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.

20

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:

21

"f. 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).

- DestroyedSemanticPred
- DirSelectRestrict
- DirectionSemanticPred
- DiscloseFactoroid
- DisclosureFactoroid
- DiscoverSemanticPred
- EmotionalStateSemanticPred
- EndSemanticPred
- EnforceSemanticPred
- EnsureSemanticPred
- EqualsSemanticPred
- ExertForceSemanticPred
- ExistSemanticPred
- ExperienceSemanticPred
- ExperienceThematicPole

GATE TEAMWARE

- Online annotation and analysis tool.
- Provides texts from a corpus and a list of elements to annotate the text. Elements could be anything, but here factors.
- Allows teams of annotators to annotate a corpus (crowdsourcing).
- Tool does automated comparison of annotations to see agreement and disagreement.
- Input to empirical investigation of legal texts.
- Development of gold standard corpora for further (machine learning) processing.

GATE TEAMWARE

Type	Set	Start	End	Id	Features
F6-Plaintiff-adopted-security-measures		2693	2917	3	

1 Annotations (1 selected) Select:

critical tolerances, dimensions, specifications and material requirements. The basic design, which played an integral role in the production of the world's first commercial jet transport, has undergone minor changes since its inception in the 1950's, each requiring FAA reauthorization.

Because these drawings are the result of Boeing's original \$ 16 million investment, and as they constitute Boeing's "lifblood" in the commercial airplane business, Boeing always considered them proprietary trade secrets. For protection of its trade secrets Boeing requires in a standard contract provision that Boeing's outside suppliers, who receive these drawings, agree not to use them for any purpose other than exclusive Boeing manufacture.

Prior FAA authorization is needed for every spare part sold or installed on any commercial aircraft. Outside suppliers may manufacture and sell airplane parts by either: (1) selling to a manufacturer with prior FAA design [*42] authorization (like Boeing); or (2) selling directly to airlines after obtaining their own FAA authorization by (a) independent design and testing; (b) licensing from another authorized airplane manufacturer (like Boeing); or (c) showing that their own drawings and manufacturing [**671] processes are identical to those previously authorized (identity). In granting an identity, the FAA does not consider where the applicant obtained its drawings or derived its manufacturing process.

The three major suppliers of aircraft windows in the United States are PPG Industries, Inc.; Swedlow, Inc.; and Sierracin. Sierracin had supplied Boeing with other products for many years. In 1982, it began manufacturing Boeing's 777 windows after acquiring the business of Libbey-Owens-Ford Company, which had previously

Annotation Editor Dialog

Type

- F1-Plaintiff-disclosed-information-in-negotiations
- F10-Secrets-Disclosed-Outsiders
- F21-Knew-Info-Confidential
- F6-Plaintiff-adopted-security-measures**

Features

Case Regexp Whole Highlights

The Boeing Co. ?

First Prev Next Annotate Ann. all next

PROSPECTS IN ONLINE LEGAL EDUCATION

- Have students and legal professionals collaborate online to identify relevant elements and passages of legal texts.
- Online annotation experiments with users.
- Creation of large, gold standard corpora.
- Empirical investigation of cases in the corpora.
- Combining standard legal education with web-based learning tools.
- Introduce additional elements to investigate.
- Distance collaboration.

THANKS!