

## INTEGRATING EBL WITH AUTOMATIC TEXT ANALYSIS

S. DELISLE, S. MATWIN, L. ZUPAN\*

Department of Computer Science, University of Ottawa, Ottawa, Ontario, K1N 6N5 Canada

\*Ecole Nationale Supérieure des Télécommunications, Paris, France

email: stan@csi.uottawa.ca

### Abstract

Existing text analysis systems (TAS) and systems for knowledge acquisition from expository texts do not process examples in a satisfactory manner, if they do at all. This is clearly a limitation since examples in texts are usually meant to show the reader how to integrate the declarative part of the text into an operational concept or procedure. On the other hand, EBL seems to fill this gap as it explains the example within the domain theory, generalizes the explanation and operationalizes the concept definition by compiling necessary knowledge from the domain theory into the definition. Not using examples to the fullest extent in knowledge acquisition from text is a drawback. Expository texts rely on examples to show to the reader how to integrate different rules contained in the declarative part of the text into an operational concept or procedure. The writers of expository texts often expect their readers to use examples to generate reasonable abstractions for using a concept or procedure.

On the other hand, Explanation-Based Learning seems to fill perfectly the gap in the existing text analysis systems. EBL links an example with the underlying domain theory by explaining the example within the domain theory, and compiling, into the example itself, parts of the domain theory necessary to operationalize the example. Moreover, the explanation and its operational part are generalized, so that the result is usable not only for the single example but for any case that has the same justification.

In this paper, we study the synergistic combination of automatic text analysis and EBL. We assume that the texts processed by TAS have the following components:

- the narrative text describing the domain
- examples, which illustrate rules, concepts, and procedures defined in the declarative part
- captions, which accompany examples and highlight their contents. Captions name the concept, or procedure, described in the example. They do not provide the details of the concept or procedure definition. Captions link the example with the narrative text.

We propose initially to use EBL on the examples, with the following mapping of parts of the expository text into EBL:

- the declarative part is converted into a domain theory. This is done by a TAS, perhaps with assistance of the human operator (e.g. in removing certain types of natural language ambiguities)
- examples of concepts or procedures from the expository text are used as EBL's training examples
- captions are used as non-operational definitions of concepts and procedures

The result of EBL, i.e. the operationalized and generalized concept or procedure, is then added to the knowledge base that has been acquired by TAS from the expository text. Although the learning that has been achieved is of the "deductive" type, it enhances the knowledge base with new and potentially useful rules that cannot be obtained from the analysis of the declarative part alone. Full account of our work describes a prototype system for INtegration of Text analysis with ExpLanation-based LeArning (INTELLA)<sup>1</sup> and shows its application on the text of the Canadian Income Tax Guide. We also discuss the necessary extensions of EBL's concept of operationality and show INTELLA's preliminary architecture. This prototype Prolog-based system consists of four principal components: NL Simplifier, Parser, Case-to-Fact Transducer, and the EBG module.

---

<sup>1</sup> Integrating EBL with Automatic Text Analysis, TR-90-41, University of Ottawa, October 1990