

Paper Review Form

Reviewer Name: Charulakshmi Vijayagopal

Paper Name: SOFIE: A Self-Organizing Framework for Information Extraction

Section I. Overview

A. Reader Interest

1. Which category describes this manuscript?
 Practice/Application/Case Study/Experience Report
 Research/Technology
 Survey/Tutorial/How-To

B. Content

1. Please explain how this manuscript advances this field of research and/or contributes something new to the literature.

This paper extends the existing technology for Information Extraction by solving problems like entity disambiguation, pattern selection and consistency checking using a single approach. SOFIE's output is canonicalized and can be used directly in ontology. SOFIE can also handle web-pages and natural language text as against the other tools which can handle information like info-boxes in Wikipedia.

C. Presentation

1. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on?
 Yes
 Could be improved
 No
2. How would you rate the organization of the manuscript? Is it focused? Is the length appropriate for the topic?
 Satisfactory
 Could be improved
 Poor
3. Please rate and comment on the readability of this manuscript.
 Easy to read
 Readable - but requires some effort to understand
 Difficult to read and understand
 Unreadable

Section II. Evaluation

Please rate the manuscript. Explain your choice.

- Award Quality
 Excellent

___ Good
___ Fair
___ Poor

Section III. Detailed Comments (provide your thoughts/criticism about the ideas in the paper; not only summarize the paper but have a critical look here)

Providing a single solution to three problems in creating ontology by extracting meaningful relationships from not just structured info-boxes but also natural language text and web-pages is brilliant. I really like the idea of weighing the facts and relationship inferences against hypotheses instead of just ignoring the hypotheses as false. The method for tie-breaking (using rules and hypotheses) between the extracted relationships of equal weights using Ockham Razor's principle is also fair. The improvements made to the existing method are really good, but the testing environment and implementation details are still unclear.

Additional Comments:

1. Provide one aspect that you liked the most in this paper.
SOFIE can handle structured text, natural language text and web-pages with fairly the same level of performance and it can be directly used in a formal ontology unlike those that already exist.
2. Provide one aspect that you disliked the most in this paper.
The implementation and the math involved in SOFIE is quiet complicated but a robust implementation is inherently complex.

Section IV. Discussion Points (provide at least 3 discussion topics/questions related to ideas/techniques described in the paper; these will be used for discussions in the class)

1. Is the test data involved in testing SOFIE fair one?
2. Is Wikipedia the main source of test data? Wikipedia is quiet recent and even the information other than info-boxes could be tagged or be slightly structured. Will that favor SOFIE?
3. What are the special cases when MAX-SAT problem is polynomial time and why are they not so common?