OVERALL EVALUATION

Rankings: [1 = unacceptable; 10 = excellent]
(do not mark those that do not apply)
{to enter your rankings in tables use the lock button from the Toolbar of Forms in MS-Word and save at the end)

Recommendations for review

My overall recommendation is:

Outstanding	☐
Definitely accept	☐
Minor revisions required before accept	☐
Major revisions required before accept	☐
Definitely reject	☒

Editorial Decision

Accept	☐
Accept after specified revision	☐
Reject	☐

Evaluation *Criteria*

Popularity of the subject	8
Appropriateness for journal	7
Adequacy of literature review	6
Quality of research design	4
Adequacy of data analysis	7
Contributions to the literature	4
Legitimacy of conclusions	4
Practical/Managerial significance	6
Clarity of presentation	7

GENERAL POINTS

A. Briefly summarize the paper in your own words, making sure to identify the paper's contribution to the field of SW and IS.

The authors present a collection of mostly syntactic methods, and a proof-of-concept implementation aiming to annotate web pages with additional information, mostly about instances mentioned in that web page. A crawler, syntactic parser, semi-automatic tagging method, and user preferences handling methods are presented.

*B How **SIGNIFICANT** is this paper?*
(How important is the problem studied? Does the paper stimulate discussion of important issues or alternative points of view? Please consider relevance to both CS and IS communities.)

☐ Very significant   ☐ Moderately significant   ☒ Not significant

Please explain your rating?

Contributions are marginal both from the theoretical and practical points of view. Concerning the theory, no new original methods are presented; the core method is just an application of Inductive Logic Programming which additionally is human-assisted.

*C. How **ORIGINAL** is this paper?*
(Are the problems and approaches new? Is this a novel combination of existing techniques? Does the paper point out differences from related research? Does it address a new problem or one that has not been studies in depth? Does it introduce an interesting research paradigm? Does it introduce an idea that appears promising or might stimulate others to develop promising alternatives?)

☐ Very original   ☒ moderately original   ☐ Not original

Please explain your rating?

Originality is restricted to a new application (web page instances annotations) of well-known methods. This applies to all of the stages of the proposed process, from the crawling to the storing of annotations, including the parsing, user preferences handling, etc.

Further, discussion of related literature is incomplete; in the "related work" section, authors seem to be completely unaware of a long tradition of approximate semantics in text corpus, like the Latent Semantics methods, and many, many others following. These are truly semantic methods, but in that section authors discuss mainly syntactic issues like crawlers and Natural Language parsers. Even the critic of Semantic Web on the grounds of wasting the current "non-semantic" web is not new, but authors don't seem to be aware of this.

*D. Please comment on the **QUALITY** of this paper?*

(Is the paper technically sound? Does it carefully evaluate the strengths and limitations of its contributions? Are its claims backed up? Does the paper offer a new form of evidence in support of or against a well-known technique? Does the paper back up a theoretical idea already in the literature with experimental evidence? Does it offer a theoretical analysis of prior experimental results?)

Experimental methodology is weak, as it relies on human work (for instance, for tagging the "injured" predicate), but no elimination methods for human bias (such as control groups) are considered. Further, selection of "adequate" rules, in this case the last 2 in table 3, is done manually. Data set in experiments is much too small for statistical purposes (38 examples); experiments are limited to measure performance of ILP rules, which is a small part of the proposed method. Generality of methods is undermined by the fact that specific web page structures are required, like pages containing tables.

*E. Please comment on the **CLARITY** of this paper?*

(Is the paper well written? Does it motivate the research? Are the % results described and evaluated? Is the paper organized in a logical fashion? Is the paper written in a manner that makes it accessible to most educators and/or educational system developers? Is the paper written in clear English? Is the readability good, average or poor? Are there any presentation problems?)

The paper is moderately well written. There are lots of English errors, missing references, etc (see list below). There are also some structural flaws, like for instance, that implementation details of the system are presented before a detailed presentation of the basic "semantization" methods.

F. Please comment on the appropriateness of the title, abstract and conclusion?

The title is a big overstatement, as authors provide an extremely limited semantic enhancement, which is just to annotate instances. A title such as "Web Semantization" suggests a much more general, ambitious –and ultimately semantic proposal.

Claims in the conclusion are too vague to be even confirmed or rejected, for instance, authors say they have "developed the idea of web semantization", claim that lacks objectivity.
Also, soft methods are only used for handling conflicting user goals, not for any semantic-related task. This limitation is not evident in the abstract and introduction.

COMMENTS TO THE AUTHOR:

i) Weak Points of the Paper:

Please drop such a title as "web semantization", as it is an overstatement, which also gives an idea unrelated to your actual work.
Experimental methodology should be revised, with particular attention to human intervention, because human-introduced bias should be avoided.

ii) Strong Points of the paper:

A complete system, representative of the state-of-the-art in many aspects has been completed.

Reviewer: Please feel free to include relevant publications and literature the authors should study or cite as appropriate in sections iii and Iv. Feel free to include URLs from citeseer, ACM or IEEE digital libraries.

iii) Recommended minor revisions? (if any)

- English writing is unnatural; should be thoroughly reviewed by a native English speaker; for instance, in the phrase "Our main contribution is the proof of concept of the idea of semantization of the web of today..." there are 5 instances of "of". Further, there are many mistakes, such as "Second idea is to split annotation process to two parts,", where the article "The" is missing at the beginning, and also the preposition for "split" should be "into"; actually, in many phrases the initial "The" article is missing. Another example is the phrase "Thanks the fact that we have...", page 8. Another one is in page 16: "Annotation says that on URL is a statement..."
- Some statements are not justified, as "Using traditional search engines has limited capabilities."
- Reference missing in phrase "A step in direction of semantic enrichment is presented also by indexes of various search engines – they represent a certain sort of semantic enrichment", page 4.
- Reference missing with phrase "the main problem of current tools with own enrichment of web content is that this enrichment mostly serves query answering and does not produce annotation of sources accessible for public for further processing."
- The set of questions just before section 3 are too different in generality, and don't point out the core problems to be solved.
- A "tectogrammatical structure" captures semantic meaning? (page 5). This seems like a confusion between syntax and semantics.

- Define "Egothor" before making statements about it (page 6).
- Reference to ILP missing, page 11.
- Please explain what are in this case positive and negative examples (page 11).
- In section 5.2, authors don't explain whether the method based on Czech linguistics is a "second approach" alternative to the previous one, or complementary to it.
- Statement "These rules represent common structural patterns that occur in sentences (more precisely in corresponding trees) with the same or similar meaning." is not justified, and further, no criterion for determining meaning similarity is provided.
- In the beginning of the experiments section, authors fail to present what is the general goal of the experiments.
- The title "Semantic Repository" (section 7) for the base produced with proposed annotations is a huge overstatement for mostly syntactic enhancements. The same for the "semantic search engine", which doesn't justify the "semantic" adjective. Actually, there are some truly semantic repositories and semantic search engines, but not the ones presented here.

iv) Recommended major Revisions? (if any)
(In case of major revisions, the revised paper will be reviewed again and authors will be asked to provide a short summary of the performed revisions along with the revised paper.)

Overall methodology, and even title, should be revised, as explained above.

How confident are you in your appropriateness as a referee for this paper?
☐ Very confident - I am an expert in this area.
☒ Confident - I have a reasonable knowledge of this area.
☐ Fairly confident - I have some knowledge of this area.
☐ Not confident - I have no significant knowledge of this area.

COMMENTS TO THE EDITOR:

This portion of the rating form is for journal editors only. Please give your reasons for your evaluations and specific criticisms.

I am completely sure about my rejection suggestion. With minor originality contribution, and several major flaws, I think it's better to plainly reject than to ask author to modify.