

Supervisor's statement of a final thesis

Czech Technical University in Prague

Faculty of Information Technology

Student: Bc. Lukáš Polák
Supervisor: Nicholas Webb, Ph.D.
Thesis title: Analyzing Supreme Court Oral Arguments with Natural Language Processing
Branch of the study: Knowledge Engineering

Date: 26. 1. 2017

<p><i>Evaluation criterion:</i></p> <p>1. Difficulty and other comments on the assignment</p> <p><i>Criteria description:</i> Characterize this final thesis in detail and its relationships to previous or current projects. Comment what is difficult about this thesis (in case of a more difficult thesis, you may overlook some shortcomings that you would not in case of an easy assignment, and on the contrary, with an easy assignment those shortcomings should be evaluated more strictly.)</p> <p><i>Comments:</i> The task undertaken by Lukas in this thesis is a challenging one. Analysis of oral arguments is difficult because the structure of spoken language often violates known or understood grammatical rules, and the majority of language processing resources used today are not trained or tuned to spoken language. These tools have for the most part been developed on structured language resources such as newswire text. To combat this, we often have to fall back on simple, word-based techniques which inevitably have lower accuracy than deeper understanding mechanisms. In this thesis, Lukas attempts to validate a number of such approaches for understanding structure, semantics and pragmatics, including polarity or sentiment, using for the most part simple word-based techniques, and to do so by analyzing and predicting the outcome of the oral arguments of the US Supreme Court. The complexity of this task is underscored by the relatively few overall attempts to do this, despite the wealth of material (the digital versions of these oral exchanges) available from the Supreme Court website.</p>	<p><i>The evaluation scale: 1 to 5.</i></p> <p>1 = extremely challenging assignment, 2 = rather difficult assignment, 3 = assignment of average difficulty, 4 = easier, but still sufficient assignment, 5 = insufficient assignment</p>
<p><i>Evaluation criterion:</i></p> <p>2. Fulfilment of the assignment</p> <p><i>Criteria description:</i> Assess whether the thesis meets the assignment statement. In Comments indicate parts of the assignment that have not been fulfilled, completely or partially, or extensions of the thesis beyond the original assignment. If the assignment was not completely fulfilled, try to assess the importance, impact, and possibly also the reason of the insufficiencies.</p> <p><i>Comments:</i> This thesis definitely meets the stated goals in my opinion. In order to do so, Lukas had to evaluate and deploy a number of existing techniques, but over new data and in a new context. This represents an implicit extension of these techniques and the findings are interesting for researchers who are interested in material beyond written prose. Ultimately, the thesis demonstrates the applicability of these 'simple' NLP techniques to discover and determine sentiment in dialogue.</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p>1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled</p>
<p><i>Evaluation criterion:</i></p> <p>3. Size of the main written part</p> <p><i>Criteria description:</i> Evaluate the adequacy of the extent of the final thesis, considering its content and the size of the written part, i.e. that all parts of the thesis are rich on information and the text does not contain unnecessary parts.</p> <p><i>Comments:</i> In fact I would say this thesis exceeds the criteria. When discussing the outline of the proposal, I think we both thought that the thesis would be shorter in length. In order to fairly represent the tasks undertaken, and to provide the requisite background knowledge necessary to explain the thesis and the decisions taken, the length of the thesis grew significantly. However, I am very pleased with the result, and feel that the information is well presented, well organized and fairly reflects the effort Lukas put into this thesis as a whole.</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p>1 = meets the criteria, 2 = meets the criteria with minor objections, 3 = meets the criteria with major objections, 4 = does not meet the criteria</p>
<p><i>Evaluation criterion:</i></p> <p>4. Factual and logical level of the thesis</p> <p><i>Criteria description:</i> Assess whether the thesis is correct as to the facts or if there are factual errors and inaccuracies. Evaluate further the logical structure of the thesis, links among the chapters, and the comprehensibility of the text for a reader.</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>92 (A)</p>

Comments:

I find this thesis to be factually correct. I think the structure of the thesis lays out the central arguments, then addresses those arguments with experiments, and discusses the results appropriately. Lukas worked hard to make the concepts in this thesis, particularly issues like the nature of the Supreme Court, clear to readers who may be familiar with them in only a casual manner. I feel that the results of this thesis therefore should be clear to the reader, no matter their level of background in NLP.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

5. Formal level of the thesis

90 (A)

Criteria description:

Assess the correctness of formalisms used in the thesis, the typographical and linguistic aspects, see Dean's Directive No. 12/2014, Article 3.

Comments:

The thesis is laid out very well, and as far as I can see adheres to the layout as dictated by the template. There is a lot of information to present, and this thesis does so using appropriate images, tables and figures. These are explained in the text. The language of the thesis, being in English, is extremely good for a non-native speaker, and I commend Lukas on presenting information so clearly.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

6. Bibliography

95 (A)

Criteria description:

Evaluate the student's activity in acquisition and use of studying materials in his thesis. Characterize the choice of the sources. Discuss whether the student used all relevant sources, or whether he tried to solve problems that were already solved. Verify that all elements taken from other sources are properly differentiated from his own results and contributions. Comment if there was a possible violation of the citation ethics and if the bibliographical references are complete and in compliance with citation standards.

Comments:

I find this aspect of the thesis to be particularly revealing of Lukas' nature. While we certainly discussed papers, indeed one of my own publications formed the basis of the ideas explored here, the majority of this bibliography was brought in by Lukas himself, independently. In doing so, he has covered and surveyed a number of technologies, consistently finding resources and references himself, addressing the state of the art in those technologies. He makes it clear where techniques are leveraged from these approaches, justifying either their use, or non-use appropriately. Consequently, I find this to be a very complete and accurate bibliography.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

7. Evaluation of results, publication outputs and awards

90 (A)

Criteria description:

Comment on the achieved level of major results of the thesis and indicate whether the main results of the thesis extend published state-of-the-art results and/or bring completely new findings. Assess the quality and functionality of hardware or software solutions. Alternatively, evaluate whether the software or source code that was not created by the student himself was used in accordance with the license terms and copyright. Comment on possible publication output or awards related to the thesis.

Comments:

I think the results of this thesis are illuminating. In a world which is increasingly focused on deeper, richer linguistic structure, it is interesting to see how powerful so called 'simple' or naive approaches are, when applied in combination. Determining the outcome of Supreme Court decisions is a task which has significant societal impact, particularly in this new administration. More generally, to be able to apply a pipeline of known techniques, tuned for the specific challenges of spoken as opposed to written discourse, and determine their validity is an accomplished result. I would be happy to push Lukas to derive one or more publications from this thesis, and feel that there are a number of appropriate venues (such as SigDial, ACL, LREC, and journals such as JNLE). The software used is used appropriately, and cited and sourced as one would expect.

Evaluation criterion:

No evaluation scale.

8. Applicability of the results

Criteria description:

Indicate the potential of using the results of the thesis in practice.

Comments:

As highlighted previously, there is a concentration in the NLP community of using increasingly complex methods (often reliant on vast amounts of textual data). While in a number of important application domains these techniques achieve state of the art performance, it is often at the cost of simplicity - both in terms of execution and in explainability of results. In contrast, the results of this thesis demonstrate that a pipeline of simple techniques can still perform extremely well in determining a complex result. And, as the methods are well understood, they can be used to explain the results found, something complex, opaque methods are unable to do. To take well understood principles and techniques, and apply them to significantly more complex, structurally different data (from prose to spoken dialogue) only underlines that while we might not have the exact linguistic structure of language correct in our tools, we are encoding elements which are at least highly correlated with the meaning in our tools, and that such correlations are transferable across language uses. This is a significant and extremely interesting (not to say encouraging) result.

Evaluation criterion:

The evaluation scale: 1 to 5.

9. Activity and self-reliance of the student

9a:
1 = excellent activity,
2 = very good activity,
3 = average activity,
4 = weaker, but still sufficient activity,
5 = insufficient activity

9b:
1 = excellent self-reliance,
2 = very good self-reliance,
3 = average self-reliance,
4 = weaker, but still sufficient self-reliance,
5 = insufficient self-reliance.

Criteria description:

Review student's activity while working on this final thesis, student's punctuality when meeting the deadlines and consulting continuously and also, student's preparedness for these consultations. Furthermore, review student's independency.

Comments:

Lukas was excellent in terms of self reliance. We met almost every week via Skype, and in between he was self motivating and self driven. He set and met his own goals, which took him in a variety of interesting directions. My role was really just as sounding board in terms of some decisions made. Where I list his activity as very good, it is only not excellent because of his other commitments. That Lukas completed this thesis while starting work is an impressive feat. Inevitably there were some weeks where progress was not as complete as it could have been, but he still worked extremely hard, and I think the quality of the final thesis is a reflection of the amount of work, both in terms of consideration application and effort, that Lukas put in over this process.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

10. The overall evaluation

92 (A)

Criteria description:

Summarize the parts of the thesis that had major impact on your evaluation. The overall evaluation **does not** have to be the arithmetic mean or any other formula with the values from the previous evaluation criteria 1 to 9.

Comments:

Over the process of supervising this thesis, I have been consistently impressed with Lukas' drive and achievement. The thesis set out what I thought was an ambitious goal, and I am extremely pleased with the results - both in terms of the technological evaluation and the conceptual handling and presentation of the topic. I find this to be a very well written and presented thesis, and I commend Lukas for his hard work on the topic. I think this has been an enjoyable and productive process for both Lukas and myself, and as a supervisor I think that is the mark of great work. This is a very good piece of work, and I think has served as an excellent learning process in many aspects of language processing for Lukas as he researched, explored and deployed extensive elements of language technology. In all regards, I think this thesis has served it's purpose, and I am delighted to have been part of the process.

Signature of the supervisor: