Supervisor's statement of a final thesis

Czech Technical University in Prague **Faculty of Information Technology**

Student:	Bc. Jan Hnízdil
Supervisor:	Ing. Jan Šedivý, CSc.
Thesis title:	Text signals relevance improvement for full text serch
Branch of the study:	Knowledge Engineering (Master)

Date: 19. 5. 2015

Evaluation criterion:	The evaluation scale: 1 to 5.
1. Difficulty and other comments on the assignment	**1 = extremely challenging assignment,** *2 = rather difficult assignment,* *3 = assignment of average difficulty,* *4 = easier, but still sufficient assignment,* *5 = insufficient assignment*

Criteria description:
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.)

Comments:
The submitted diploma work researches the optimal selection of textual document signals for the purpose of full text search. The signals selection is the key for sucessfull search results ranking. The user behaviour and demands are slowly changing and the search engines need to answer more challenging queries. It is moving from a simple text search in a direction of understanding the users needs. This has to be reflected in a search of new, more appropriate signals enabling the new search styles. In this respect the work is innovative and contributes to the latest trends in IR.

Evaluation criterion:	The evaluation scale: 1 to 4.
2. Fulfilment of the assignment	**1 = assignment fulfilled,** *2 = assignment fulfilled with minor objections,* *3 = assignment fulfilled with major objections,* *4 = assignment not fulfilled*

Criteria description:
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.

Comments:
The assignment was fulfilled. The theses is suggesting new textual signals and experimentally estimates their discriminative power compared to traditionally used set of signals.`in addition some of the future direction of interesting research are suggested.

Evaluation criterion:	The evaluation scale: 1 to 4.
3. Size of the main written part	**1 = meets the criteria,** *2 = meets the criteria with minor objections,* *3 = meets the criteria with major objections,* *4 = does not meet the criteria*

Criteria description:
Compare the size of the written part with the expected size (without appendices), see the Dean's Directive No. 9/2011, Article 3. To evaluate the thesis it is also important that all parts of the written part are rich on information and necessary for a final thesis. The text should not contain unnecessary parts.

Comments:
The diploma work has over seventy pages, which is slightly above the average. The central part is the Chapter 4, Experiments. It is 27 pages long and lists, explained and discusses all the experiments. The Chapter Conclusion summaries the key conclusions and deliverables.

Evaluation criterion:	The evaluation scale: 0 to 100 points (grade A to F).
4. Factual and logical level of the thesis	*99 (A)*

Criteria description:
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.

Comments:
The whole diploma work is written in a logical sequence, explaining particular topics from general to details. The chapters are informative and concise too. The text is reasonably easy to read. The Appendix A contains the acronyms explanation. The chapters have the usual order logical order. The text is accompanied with tables, graphs and figures giving the reader deep insight to the topic.

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 aspect s, see Dean's Directive No. 9/2011, Article 3.

Comments:

The whole diploma work is carefully written in a style appropriate for a technical document. The figures, tables, equations and graphs are correctly numbered with attached explaining captures. Also the graphical appearance is very pleasing, which is not usual in technical documentations. It is written in English, which is almost a must for projects of this type. However, Jan has made some minor gramatical mistake, which are not impacting understanding.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
6. Bibliography	99 (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:

The diploma work is presented in correct formula language. Graphically and typographically is well prepared. The bibliography is organised in a style usual for the scientific documents. Throughout the text all statements are supported with references to literature or proved by an experiment. He is listing the basic most relevant references.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
7. Evaluation of results, publication outputs and awards	99 (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:

Jan has fully exploited the recommended bibliography and has aded many interesting paper references. All statements are either documented with literature references in a usual manner. The results of his work are mainly displayed in a graphical or tabular way, since these are the outcomes on numerous experiments. The results are well documented and discussed.

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:

Jan was researching new text signals for an information retrieval system. Seznam.cz offered for the diploma development a database of documents signals along with related queries. We knew which of the signals are text related. The project task was to develop new more relevant signals using the available data for testing. The diploma work presents a set of new features and it shows its impact on the relevance ranking on SERP. This results will be used for further research of the IR system in Seznam. We are almost sure, that some of the signals will enhance the current signal vectors in the product. The set of the new text signals is a further step in the development of the IR ranking.

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:

Jan was very active long before the diploma theses was formulated. He started to work before the summer 2014. We had regular meetings once a week to discuss further steps, progress and set new targets. He always came prepared with well documented progress in a shared file. He was suggesting his own ideas how to proceed.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
10. The overall evaluation	99 (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:

Jan is definately one of the best students I have met. He has worked systematically over almost one year. He has continuously discussed the steps with me and with mentors from Seznam.cz. He has put together very interesting results for improving the search results. the final experiments results is the major contribute of his work.

Signature of the supervisor: