Brno University of Technology
Faculty of Information Technology

Review of Bachelor's Thesis

Student: Khudiakov Daniil
Title: Corpus Processing for Foreign Language Learning (id 18763)
Reviewer: Dytrych Jaroslav, Ing., UPGM FIT VUT

1. Assignment complexity — **less demanding assignment**
The student had an opportunity to handle the assignment as less demanding or very demanding. His approach was in effect less demanding.

2. Completeness of assignment requirements — **acceptable under serious reservation**
Assignment requirements were met, albeit in a minimal way.

3. Length of technical report — **within minimum requirements**
The technical report has 27.6 standard pages and three pages of images. Thus the minimal extent requirement was met, yet the description of student's own contribution is inadequate.

4. Presentation level of technical report — **50 p. (E)**
The technical report does not follow a clear logical structure. Description of development activities performed by the student is unclear and confusing. Many tools are described in the introductory theoretical part without proper discussion of merits. There is a lot of information which is not used or followed upon in other parts of the thesis. The design chapter starts with a description of requirements and a rather theoretical discussion. Design itself is covered by only three pages of text while crucial information is missing. Applied technologies aren't sufficiently described (only the search engine and one other library are described). Basis frameworks whereupon the solution is built aren't described at all. The implementation description part then partially complements the missing design information, yet it also contains too much of low level implementation detail, as for example the analyzer configuration. The evaluation of the final solution is not clear and for instance table 6.1 should be better explained and discussed.

5. Formal aspects of technical report — **55 p. (E)**
There are many neglignece errors (typos and spelling) in the thesis and some typographic violations such as one letter words at the end of a line. Czech abstract is poorly translated and some keywords were translated incorrectly. Some grammar mistakes are also in the English text. Some abbreviations are wrongly expanded at some places.

6. Literature usage — **48 p. (F)**
Crediting of third party resources is unclear and confusing. For example picture 4.2 can be found in materials by Jan Pomikálek from FI MUNI yet the source is not referenced. In 2.1.5, a resource is cited that doesn't contain the cited information (it contains only a reference to the actual source). Bibliographgy doesn't conform to the required norm and particular entries don't have uniform format (wrong usage of the template and inconsistent information in the note entry). Some easily accessible information, such as author names is sometimes missing. Additionally, many online resources are mentioned in the text without URLs which could have been for instance in footnotes.

7. Implementation results — **50 p. (E)**
The extent of the actually developed solution is so minimal that it is on the edge of acceptability. Several features, some of them rather basic, are mentioned just as possible future improvements. Usage of the relational database is ineffective. Some server functionality should have been developed as easily pluggable modules (e.g. the machine translation service, which is by the way of a rather poor quality). There are only very few comments in the code and documentation of the code is insufficient.

8. Utilizability of results
The resulting product is available online and can be publicly used. Usage statistics is not available.

9. Questions for defence
- Why are the to employed frameworks, technologies and their integration given so little space in the technical report ?

10. Total assessment — **50 p. sufficient (E)**
The extent of the actually developed solution is on the edge of acceptability. The technical report has unclear logical structure and the final solution is insufficiently described. Tools are described without discussion and/or comparison and it seems that the student mainly presented results of previous studies which aren't always cited properly. The theoretical explanations continue in the design part of the thesis and the design description itself is concluded in the implementation part. There are also some formal and language mistakes. Nevertheless,

the resulting product is usable and publicly available.

In Brno 1. June 2016

....
signature