



Course report – Faculty of Librarianship, Information, Education and IT

Name of course: Information retrieval for digital libraries 2	LADOK code: NLID23
Number of higher education credits: 7.5 credits	Period (e.g. P1 2018): P1-2 autumn 2018
Within study programme or alt. single-subject course: Master's programme: Library and Information Science, Digital Library and Information Services (BMDD1)	

Comments about response rate for the student evaluation

9 out of 24 registered students (38%) have answered the course evaluation survey.

Analysis of:

Student results and student performance on the course

24 students were registered on the course. At the end of the course period, 15 students (62%) had finished all course components. The grades for the entire course were distributed as follows:

Grade A: 3 students
Grade B: 8 students
Grade C: 4 students
Grade D: 0 students
Grade E: 0 students

The course is divided into two course components: *Assignment: Group assignment*, which is a practical text categorization task using the R language, and *Take home examination*. The group assignment was performed in work groups consisting of 4 students each, and the individual contributions were recorded in a logbook. 5 group reports were submitted for the first examination round and all reports passed. Judging from individual records in the logbooks and the extensive details put into the group reports it can be concluded that this examination task successfully invites reflection on methodology and scientific interpretation. It is also noticeable that a relatively high proportion of the active students have finished the course within the allocated course period, all obtaining a final grade in the interval A-C.

How the course contents, teaching and examination have supported student learning

58% of the respondents answer “Broadly agree” or “Completely agree” to the question pertaining to whether the teaching on the course has supported their learning. In addition, 67% of the respondents have answered “Broadly agree” or “Completely agree” to the question whether the course examinations have been relevant in relation to the learning objectives of the course. An approach that was tested on this course in order to prepare for the residential period, and to assist the learning of the course content, was to use the message board to publish short essays on various topics. In addition, a couple of exercises aimed to be performed in R were published, with the aim to gradually increase the understanding of how this tool can be used for text categorization. This seems to have been well received and compensated to some extent in the case the student could not participate in the residential period.

A problem that is highlighted in the course evaluation is the fact that this course partly runs in parallel with the Research Methods course, and that these two courses ought to be better coordinated with regard to teaching and examination tasks. Another issue raised in the evaluation is that the group assignment is perhaps not necessary on this level and that it might affect both the learning rate as well as the final grade since the task is perceived to be difficult to coordinate within the work group. There is also a concern expressed that the performance of the group will affect the final grade on the course. However, 66% of the respondents answer “Broadly agree” or “Completely agree” on the question as to whether the examination modes supported their learning. The specification of the group task was also perceived as unclear and would have needed stricter guidelines. A reflection made by the involved teachers is also that the dataset is limited with regard to its extent and that the categories involved are imbalanced. Assuming that the scope and focus of the assignment is retained, a revised or an entirely new dataset should be considered. The overall impression of the viewpoints stated in the course evaluation is therefore that the examination tasks contribute pedagogically to the understanding of the subject area, but that the group task needs elaboration in order to clarify its inherent expectations.

How the course has been linked to research

Formally speaking, the course is designed to link both theoretically and empirical to text categorization research, with a focus on how to perform and improve upon the performance of various text categorization experiments. However, only 33% of the respondents answer “Broadly agree”, and 0% answer “Completely agree”, to the question concerning whether they have gained an increased understanding of research. In addition, 22% of the respondents answer “Do not agree at all” and 56% answer “Partly agree” to the question concerning whether they have developed their academic approach. Also, 44% of the respondents answer “Partly agree” to the question concerning whether the students have developed their critical thinking through the course. On the other hand, 89% of the respondents answer “Broadly agree” or “Completely agree” on the question regarding whether they have gained an increased knowledge of research within the disciplinary domain of the course. These figures together strongly indicate that the course needs a certain reassessment of how the teaching, and not least the examination tasks, should be conducted in order to strengthen the focus on research processes within automatic text categorization. A tentative conclusion is that the course provides a theoretical understanding of the research processes associated with automatic text categorization, but that more effort needs to be put on how to convey a practical know-how with regard to formulating suitable research questions, as well as selecting and possibly even implementing practical methods and tools for investigating the aforementioned research questions. In a comment to one of the survey questions, it is hinted that the IR courses should be placed side by side (or in succession) in the same semester in order to accomplish a synergy and better equip students with an understanding of how IR research is and can be performed.

How the course has worked in relation to other courses in the study programme

88% of the respondents have answered “Partly agree” or “Broadly agree” to the question about whether it was clear how the course fits into the programme. In a comment to this question, it is pointed out that the two IR courses have been the most rewarding courses on the programme, but that these two courses are “isolated” from the other courses on the programme. This makes it difficult to relate the content of the IR courses to the programme as a whole. It is pointed out that it would be better if also other courses would be partly oriented toward computer science and IR in order to establish a foundation and connect better to each other.

How the course resources have been utilised and experienced

Most of the teaching on the course has been concentrated to the residential period, which took place a few weeks into the course period. Three teachers have been involved in lectures and practical exercises in R. A recorded lecture summarizing important aspects of the course content was published relatively late in the course period. Whether this particular lecture was helpful or not for the preparation for the home exam is not indicated in the course evaluation. A problem touched upon in the evaluation is that not all students had the opportunity to participate in the teaching during the residential period. However, it seems that the amount of information provided in the course module (not least on the message board) as well as the R exercises published in advance have to a certain extent compensated for this problem. A discussion forum has been provided but was not utilized by the students. In a comment on one of the questions in the course evaluation, it is hinted that the students felt that the teachers were present and accessible, but that they had no particular need to contact them. However, it is also indicated that the students would have needed additional support concerning the group task.

Concerning the course literature, it is stated in the course evaluation that the literature was useful but that the students found themselves consulting online material as well in order to get a better understanding of various concepts. 77% of the respondents answer “Broadly agree” or “Completely agree” on the question concerning whether the course literature and other course materials supported their learning. It is also pointed out in the comments that the course book Modern Information Retrieval presents IR theory in a difficult way which makes it necessary to coordinate the reading of the book with online searching to verify that they have understood things correctly. A reflection made by the involved teachers is that also the chapters on text categorization would benefit from a study guide similar to the one that was prepared for the first IR course (Information retrieval 1).

Summary and additional comments

The overall impression of the results on the course, as well as the input provided in the course evaluation, is that it has a design that works quite well to reach the learning outcomes and that it provides the students with a theoretical and practical basis for understanding the fundamental problems and concepts of text categorization, as well as how to perform basic text categorization research. It also utilizes a platform (R and RStudio) that is commonly used both in the academia as well as in industry for statistical tasks in general and text mining in particular. However, the evaluation result also indicates that more effort needs to be put on conveying how text categorization is researched and what research problems that are investigated. An observation made by the involved teachers is also that semantic modelling is an important component on the course that currently is somewhat underdeveloped and needs to be introduced more systematically in the teaching material. The examination tasks fit well with the pedagogical objectives of the course, but the group task needs to be revised and clarified. It has also been pointed out that the course literature is difficult and needs to be complemented by study guides and possibly other kinds of material that can clarify the key concepts of the course.

Any proposed changes

Last year (2017) a major change of was made with regard to the group assignment. Instead of using the WEKA platform, it was decided that the statistical language and platform R should be applied. In conjunction with this change it was also proposed that the group assignment should focus on a hitherto ill-recognized problem, namely cultural heritage as subject material for digital libraries. As noted in the previous course report (from 2017), the revised group assignment was received as inspiring and the generally high quality of the reports produced was also an indication of a positive direction for this examination task. However, in this year's course evaluation it was pointed out that the requirement to work with the examination tasks in small work groups acted as an impediment rather as a helpful work form. It is also stated that the specification of the group task was unclear and difficult to follow without further instruction. The involved teachers have also noticed that the subtask involving semantic modelling needs more theoretical and practical preparations to better integrate with the overall research objectives of the group assignment.

Based on the observations above, we propose that the group assignment is turned into an individual assignment. The focus on a practical text categorization task should be retained, but a different dataset should be considered and the instructions need to be improved. A guide containing study questions for the course literature should also be developed, similar to the material that is being used on first IR course (Information retrieval 1).

Course Coordinator:

Johan Eklund