Course report – Information Retrieval for Digital Libraries 2 (NLID23), 7.5 credits, fall 2017 (HT-17)

The course is offered in the Master's study programme in Library and Information Science: Digital Libraries and Information Services (MADL). It has been going on during the second half of the fall term 2017, from November to January 2018. When the course started, according to the PingPong, 26 students were registered, with 14 of them active throughout the course.

Apart from one residential week in Borås (4-8 September, 2017), designed to cover both the IR1 and IR2 courses, teaching has mainly been conducted in distance mode. The meeting that focused on this course was three days, with 6 lectures, 5 hands-on practical sessions, plus a seminar to discuss credit-bringing teamwork on text mining as a method of advanced access to digital content.

Course content and teaching was structured in accordance with the following themes and problematics:

- Introduction to DLIR2; Text categorization; Evaluation; Feature selection; Information visualization.
- In teamwork, a complex set of related tasks about motif identification in a folktale dataset by text mining and information visualization, with the results to be submitted as an analytic report.

The above mentioned parts of the course were examined one at a time. In the first step, the teams had to submit their reports which had to follow a previously agreed upon structure and formal requirements. Team members were graded on a group basis. In the second step, students had to submit their home exams which were graded individually. The results were as follows:

- Component 1 (teamwork): 4 team submissions, one of them VG "strong/solid" = A "Excellent", 3 of them VG "Ordinary" = B "Very good";
- Component 2 (home exam): 10 submissions, 2 of them VG "strong/solid" = A "Excellent", 4 were VG "Ordinary" = B "Very good", and 4 were G "Strong/solid" = C "Good";
- With several students not having submitted their home exam yet, complete statistics for the course do not exist for the time being.

The course evaluation was conducted through a digital questionnaire published in Ping Pong at the end of the course. With "less than 3" responses two weeks after the course's end, the results could not be evaluated.

This year has seen a major reorganization of course content in two respects:

- We abandoned Weka for text mining and decided to use R instead;
- The new team task was to automatically identify motifs in folktales from an international collection of 381 tales in English, following traditional research conventions in the field.

From the quality of the home exams it became apparent that the students took their opportunity to express their professional stance as a digital librarian seriously. The following general remark

seems to be a fair summary of their opinion: *"...some questions I felt a bit uncertain about, however it's not a big issue as I've made a conscious decision to continue learning text mining/classification using both R and Python so I will be correcting myself the hard way I suppose. Thanks for the course (because the resources provided in it are excellent and will most certainly be helpful to me in the months ahead)"*

The high quality team reports showed that the students appreciated their new ability to work on topics of professional interest. At the same time, skills learned in teamwork helped them to master the home exam on their own which explains the overlap by design between the two components.

Those presenting in this course are grateful for feedback from our students and we will consider as many of the recommendations as feasible.

Borås, 18-02-11

Sándor Darányi, course responsible