

# Course report – Information Retrieval for digital libraries 1 (NLID12), 7.5 credits, fall 2017

The course is offered in the Master's study programme in Library and Information Science: Digital Libraries and Information Services (MADL). It was delivered during the first half of the fall term 2017, from August till November. When the course started, according to Ping Pong statistics, 67 students were registered, with 36 of them active throughout different parts of the course.

With a residential week in Borås on September 4-8, 2017 for both this course and DLIR2 (NLID23) to follow immediately after DLIR1, teaching was conducted in a combination of local vs. distance mode. All the arrangements for problem solving and clarifications were resolved in Ping Pong.

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

- Information representation for IR; Classical IR models; IR and the web; IR evaluation.
- Understanding the overlap between IR and new trends in knowledge organization with a bearing for concept-based IR and the Semantic Web. As abstracting skills are a key component of topic description and the ability to extract keywords for query formulation, to this end the students had to summarize Chapter 4 of "Information Retrieval and Knowledge Exploration" in Gödert, W., Hubrich, J., Nagelschmidt, M. 2014. Semantic Knowledge Representation for Information Retrieval. Göttingen: Hubert & Co. GmbH & Co. Topic 1 covered a summary of Sections 4.1-4.3 (pp. 61-71); topic 2 had to sum up Sections 4.3.1-4.4 (pp 71-86), both in an essay-like manner of fixed length text (5 pages respectively).

The above mentioned parts of the course were examined one at a time. In the first step, the students had to submit their essays which were graded separately on an individual basis and the grades were averaged for this component. In the second step, students had to submit their home examinations which were graded individually. The results were as follows:

- Component 1 (essays 1 & 2): submissions from 35 students, 16 of them VG = A (excellent), 13 VG = B (very good), 6 G = C (good), 1 submitted only one essay and opted for a future date to finish the course;
- Component 2 (home exam): 33 submissions, 12 of them VG = A (excellent); 13 VG = B (very good); 2 G = C (good); 4 G = D (satisfactory); 1 U = FX (insufficient); 1 U = F (insufficient).
- Course grades (passed): 12 VG = A (excellent); 13 VG = B (very good); 5 G = C (good); 1 G = D (satisfactory), altogether 31. Two students are scheduled to resit the home exam, 3 students plan to finish the course next year.

The course evaluation was conducted through a digital questionnaire published in Ping Pong by the end of the course. 13 of 67 registered students (13 %) have filled in the questionnaire, which meant actually 13 of the 36 active participants (36 %). Responses to the questionnaire are presented in the attachment.

The evaluation once again revealed the importance of the residential week which, for several students, is the forum to recall secondary school mathematics necessary to follow the course. With this opportunity missing for those who could not show up, a few clearly struggled with making sense of the course. Several people criticized the essay task because they did not see the link between text summarization and IR, however the majority coped with the challenge and seems to have understood its significance.

The course textbook was difficult for some but seems to have been considered to have covered the essentials for many, reflected in the overall very good course grades. At the same time thick as it is, the same textbook is becoming outdated given the fast tempo of progress in IR research. Once the increasing amount of knowledge to be covered vs. a limited time to present the basics meets the limited willingness of some to face reality, problems can be expected.

Another interesting and so far unseen complaint was regarding the English of a textbook (by German authors) or that of the course teachers. Clearly, as the course sees more and more students from the UK and the US, they introduce their own linguistic expectations. Plus on the one hand, there are always students who want to learn more – here, someone suggested to add artificial intelligence to IR –, or less, with people asking for more and more predigested and simplified knowledge e.g. about the Semantic Web, or to find even more simple ways of explanations for everyone with diverse backgrounds. Whatever one thinks, it is impossible to please everyone but university level education is for grown-ups.

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

Borås, 17-12-18

Sándor Darányi, course responsible

Attachment 1:

# Survey results

**Survey**DLIR1 course evaluation 2017

**Event**NLID12 H17-1 Information Retrieval for digital libraries 1, 7,5 credits

BMDD116h, BMDD117h

**Status**open

**Date**2017-11-22 12:17

**Group**Participants

**Answered by**13(67) (19%)

**1 What is your opinion about the structure and presentation of the course, i.e. links between lectures, literature, tasks and examination, residential period activities? Has the residential period contributed to your understanding? What do you suggest to change and why?**

---

<b>Answer choice</b>	<b>Poor</b>	<b>Non-satisfactory</b>	<b>Satisfactory</b>	<b>Good</b>	<b>Very good</b>
<b>Distribution</b>	0%	16,7%	50%	16,7%	16,7%
<b>Number</b>	0	2	6	2	2

---

12 have answered of 67 (17%)

Maximum number of choices: 1

---

## **Respondents comments:**

- The residential period really helped even though I couldn't "catch" the meaning of many things that were said as I hadn't read much of the course material by then. The notes I took from lectures though, helped me later on as I was reading the course material again and again.
- Without a background in informatics, mathematics and/or logic, the found the course structure was difficult.

The residential period did not contribute towards my understanding because there was simply too much to take in. This is understandable, but I believe that more should be done to prepare students for the residential period. In this way, I think that the residential period would be more rewarding.

The literature, while relevant, was not very pedagogical. I found myself referring to other books as well as lectures from other universities in order to grasp some fundamental concepts of IR. Perhaps this is expected of students.

---

- Everything was very confusing, I wasn't aware of the residential period so I missed it and from then all the document seemed to be in a weird order. The study guide was before the actual lessons for example.
  - I appreciated very much the material that we were given, especially the study units and the study guide. I felt actually overwhelmed and could not follow everything, mostly as a matter of time, but as I intend to study it further, it is very useful material for explaining and as use of a guide. The residential period was also more than useful, as the presentations were concrete and gave a lot of new information on new fields with simple understandable ways. I think I would not be able to follow the course if I had not participated in the residential week.
  - NO comment - does not wish to burn bridges since this course clearly lacks any sort of way to be anonymous.
- 
- Generally, the connections between the residential period, the material, and the exam were good. More information on the exam might be useful in guiding our reading, since there is quite a lot of reading. I found some of the material to be much easier to read, while some readings were quite tricky to understand. I might direct students towards the more simply written materials first, with the less clear materials to be read afterwards. The unit guides were useful and simple, as was most of the Belew book. The MIR book is useful but poorly written - it was always helpful to do other readings before looking at this text. The Saracevic article was very difficult to understand. If there is anything else on the subject, it would be good to swap out this article. In terms of the residential week, I think this was done pretty well, especially considering that IR1 and IR2 needed to share the week. Making the difference between IR1 versus IR2 lectures a little clearer might have been helpful, but I especially appreciated the focus on maths, which re-acclimated me to linear algebra, geometry, etc. as early as possible.

**2 Was the course guide and instruction material on PingPong relevant?  
Motivate, please.**

**2.1 Course guide and instruction material:**

---

<b>Answer choice</b>	<b>Poor</b>	<b>Non-satisfactory</b>	<b>Satisfactory</b>	<b>Good</b>	<b>Very good</b>
<b>Distribution</b>	7,7%	0%	46,2%	23,1%	23,1%
<b>Number</b>	1	0	6	3	3

---

13 have answered of 67 (19%)  
Maximum number of choices: 1

---

## Respondents comments:

- Instructions were clear. The step by step guidelines about what chapters to read for each unit were really appreciated.
  - I find myself coping text to a worddocument to make it easier to read. I long for an old fashion text feeling and less words.
  - The course guide and material were very relevant and easy to navigate and implement.
  - Again, it was arranged in a very confusing fashion, if you would had lecture first, then study material with title related to the lecture, that would have helped. A lot of the files' name made no sense to me and I had to click them all to see which one were relevant to what I was doing.
  - (see above)  
The course guide and the material for the study units were very useful as a guide to make a concrete idea about the area of the IR and know where to look for when needed. I appreciate very much the way it was written, simple to someone who is new to these notions, but explaining and giving also the big picture.
  - The instruction material for the home essays were appalling; change of format 10 days before submission (change from 1.0 word spacing to 1.5, thats almost 2 full pages less for someone that actually wrote 5 pages in 1.0) What was actually supposed to be written about was not clearly stated until students contacted responsible teachers for the master program rather than The course teacher, that returned snide remarks in emails which is very unprofessional. It was stated as an "essay" and simply that. If you look at the instructions in the essay topic document, or the introduction - it is a completely different assignment compared to the "clarification" that was added 10 days prior deadline.
- 
- The course guide was easy to follow.

## 2.2 Selection of literature:

### 2.2.1 General amount:

---

Answer choice	Not enough	Sufficient	Too much
Distribution	15,4%	76,9%	7,7%
Number	2	10	1

---

13 have answered of 67 (19%)  
Maximum number of choices: 1

### 2.2.2 Complexity:

---

<b>Answer choice</b>	<b>Too difficult</b>	<b>Challenging</b>	<b>Too easy</b>
<b>Distribution</b>	23,1%	76,9%	0%
<b>Number</b>	3	10	0

---

13 have answered of 67 (19%)  
Maximum number of choices: 1

### 2.3.1 Unit 1:

---

<b>Answer choice</b>	<b>Poor</b>	<b>Non-satisfactory</b>	<b>Satisfactory</b>	<b>Good</b>	<b>Very good</b>
<b>Distribution</b>	0%	0%	61,5%	15,4%	23,1%
<b>Number</b>	0	0	8	2	3

---

13 have answered of 67 (19%)  
Maximum number of choices: 1

### 2.3.2 Unit 2:

---

<b>Answer choice</b>	<b>Poor</b>	<b>Non-satisfactory</b>	<b>Satisfactory</b>	<b>Good</b>	<b>Very good</b>
<b>Distribution</b>	0%	0%	53,8%	30,8%	15,4%
<b>Number</b>	0	0	7	4	2

---

13 have answered of 67 (19%)  
Maximum number of choices: 1

### 2.3.3 Unit 3:

---

<b>Answer choice</b>	<b>Poor</b>	<b>Non-satisfactory</b>	<b>Satisfactory</b>	<b>Good</b>	<b>Very good</b>
----------------------	-------------	-------------------------	---------------------	-------------	------------------

<b>Distribution</b>	0%	0%	61,5%	23,1%	15,4%
<b>Number</b>	0	0	8	3	2

---

13 have answered of 67 (19%)  
Maximum number of choices: 1

#### 2.3.4 Unit 4:

---

<b>Answer choice</b>	<b>Poor</b>	<b>Non-satisfactory</b>	<b>Satisfactory</b>	<b>Good</b>	<b>Very good</b>
<b>Distribution</b>	0%	0%	61,5%	23,1%	15,4%
<b>Number</b>	0	0	8	3	2

---

13 have answered of 67 (19%)  
Maximum number of choices: 1

#### Comments on literature:

---

0 have answered of 67 (0%)  
Maximum number of choices: 1

---

#### Respondents comments:

- I started of reading the books and lectures but more and more I ended online to find better examples. I worry a lot that I will misunderstand questions. I do my best to study in a classical way. Book and lectures first.
- I honestly got more help from the Unit study guides than the literature itself.
- The most important and really helpful source (especially for the Home Exam) for this course was the book of Baeza-Yates and Ribeiro "Modern Information Retrieval".
- Not very pedagogical. I often used other resources as my primary ones and the textbooks as secondary resources.
- The text we had to abstract was written in weird English, with a lot of colloquialisms (for example: resp.). It would have been good if that had been addressed to help with the reading.
- I have not read all the given literature. But what I enjoyed much was the FOA book, that gives the cognitive perspective on the IR processes.
- The essays required the same amount of text submitted, while the chapters were completely different (ie the 3 last pages of essay 2 were empty. The language was not the easiest, it seems like it either was poorly proofread or we are supposed to

understand 2nd language english based on german structuring, use of words were a bit weird - would have helped if we actually had some sort of unit where it was discussed in depth.

- See answer to question 1. I was a little fuzzier on the scope of the last unit. The Semantic Web just seems like too big a topic to bookend the course. It might be difficult to find more directed literature on SKOS though. But more (or simply more clearly stated) focus in this section would have been appreciated.

---

#### **2.4 Is there a topic that you think should be studied in more detail in the course?**

---

0 have answered of 67 (0%)  
Maximum number of choices: 1

---

##### **Respondents comments:**

- I liked the assignments in math. Thank you.
  - I haven't thought of it.
- 
- Not really, though see above comment on unit 4.

#### **2.5 Is there an area or a topic that you think is missing from the course? Please explain.**

---

0 have answered of 67 (0%)  
Maximum number of choices: 1

---

##### **Respondents comments:**

- I like IT and math. Perhaps a discussion about AI and search hits in Google/Duckduckgo/more?.  
On the subject of AI:  
<https://arxiv.org/pdf/1707.02268.pdf>  
About search hits on Google:  
-What is gclid?  
<https://www.wordstream.com/gclid>  
Gclid is a globally unique tracking parameter (Google Click Identifier) used by Google to pass information back and forth between Google AdWords and Google Analytics. If you enable URL auto tagging in Google AdWords, Google will append a unique ?gclid parameter on your destination URLs at run-time. Because it is a redirect, you won't see any gclid parameters on your ad words text ad destination url's, but it will show up in your Web server log files. Auto tagging was introduced in 2004 and is on by default in any Google AdWords accounts

An unfortunate side effect of Google Adwords & Google Analytics integration is that it kills your search query data.

Critical data pertaining to clicks and conversions are mis-attributed to the keyword in your AdWords account that triggered your ad, not the actual search query that was entered by the searcher.

- 
- No.

### 3. Were the tasks (exercises) meaningful?

---

Answer choice	Poor	Non-satisfactory	Satisfactory	Meaningful	Quite meaningful
Distribution	0%	0%	61,5%	23,1%	15,4%
Number	0	0	8	3	2

---

13 have answered of 67 (19%)

Maximum number of choices: 1

**Please indicate the least relevant ones. Explain.**

---

0 have answered of 67 (0%)

Maximum number of choices: 1

---

#### Respondents comments:

- I dont know. I haven't come up with any conclusion. Still waiting for home exam.I might have misunderstood questions. I found myself searching online for answers.
  - Exercise no.3, about Page Rank was not so clear to me about what to do.
  - I thought the abstract was not very relevant as it was meant to be a summary but the length asked was almost the same as the actual text. It wasn't clear what we needed to do, it seemed to be more of a translation of a poorly written text.
  - In my opinion PageRank as a subject is more semantic - wouldnt it be better to spend that topic on theoretically discussing how PageRank values are done rather than straight up math of things that are trade secrets and deemed to change every day as it is very dynamic?
- 
- I liked the exercises, which made the readings concrete and forced us to do the calculations ourselves. Honestly, I think that they should be mandatory, though I understand why they are not.

### 4. Did you perceive your own input as sufficient for studying the course?

---

<b>Answer choice</b>	<b>Poor</b>	<b>Insufficient</b>	<b>Satisfactory</b>	<b>Sufficient</b>	<b>Quite sufficient</b>
<b>Distribution</b>	0%	16,7%	50%	33,3%	0%
<b>Number</b>	0	2	6	4	0

---

12 have answered of 67 (17%)

Maximum number of choices: 1

### 5. What do you think of the home examination's relevance and meaning?

---

<b>Answer choice</b>	<b>Poor</b>	<b>Irrelevant</b>	<b>Satisfactory</b>	<b>Relevant</b>	<b>Quite relevant</b>
<b>Distribution</b>	7,7%	0%	46,2%	15,4%	30,8%
<b>Number</b>	1	0	6	2	4

---

13 have answered of 67 (19%)

Maximum number of choices: 1

### 6. Other comments and general evaluation of the course:

---

0 have answered of 67 (0%)

Maximum number of choices: 1

---

#### Respondents comments:

- Still waiting for the result of home exam. I might have misunderstood questions. I found myself searching online for answers.
- I got a lot out of this course but due to extended family and personal illness was unable to dedicate as much time to this course as I would have liked. Although I could have gleaned much more from the material and exercises, I still consider myself improved from taking this course and feel as if I have a decent foundation with which to move forward.
- The content of this course was very interesting though it was not so easy by its nature. I needed to read the material several times in order to understand it, as I have only humanities studies background.

The home examination contained all issues of the course and it was given in such a way to help student pass. The points were fairly distributed between the questions. I did a lot of reading and I wanted just to pass but, to my surprise I got a "B" (!). A remark about the home exam: it would be better if a few more days were given because there are questions that need many hours to be answered.

The teachers were really helpful and always willing to help and answer our questions.

- The formulation of the question in the home exam were very confusing, some of them were barely in English, translating the question into something I could understand made up a big part of my time. However, once that was done, the exam was very challenging but I think I learnt a lot from it.
  - It was quite difficult for me to follow it, mostly I think because of the vast amount of new learning objects, and not having sufficient time for me to study as I wanted. But I got excited by this area of studies and work, and I would like to further include it in my studies in some way. The tasks were very meaningful and useful for understanding through practice the notions they referred to.
  - Proof-read the examination, not everyone understands swedish expressions or sentence structure translated to english. Some questions made absolutely no sense and a lot of time had to be spent just translating the questions rather than answering them. Examples = "categories of information" "short and consice" "Suppose that in documents, function words occur up to 50 %. As they are regarded stopwords, whereas not 50 % of the vocabulary consists of stopwords, this is an apparent contradiction" This is not proper english!
- 
- There were a few questions on the home exam that were tricky to understand. This is a general problem with writing about math and IR systems - there is a lot of area-specific terminology that makes clear statements and questions difficult. Still, I might try to simplify the language (not the content) of some of the questions.

Overall, I enjoyed this course and believe that I learned quite a bit about a topic with which I initially had very little familiarity.