Spikker - Sieverink, B. (Barbara, CES)

From: Hamer - Mulder, I.C. den (Ida, EWI)
Sent: donderdag 30 januari 2014 12:12
To: Padberg - Heskamp, S.G. (CES); Langerak, R. (EWI); Spikker - Sieverink, B. (Barbara, CES); Hoeven, G.F. van der (EWI); Ferreira Pires, L. (Luis, EWI)
Cc: Aly, R.B.N. (Robin, EWI); Apers, P.M.G. (EWI); Hiemstra, D. (Djoerd, EWI); Jonker, W. (EWI); Keulen, M. van (Maurice); Willem Jonker (willem.jonker@ictlabs.eu)
Subject: Vakkenwijziging leerstoel DB
Attachments: DS-concept.pdf

Aan: OLC-IT, OLC-BIT en OLC-HMI

Van: Leerstoel Databases

Hierbij ontvangt u een voorstel voor een nieuw vak "Data Science" ter vervanging van "Data Warehousing en Data Mining" en "XML & Databases 1".

Wij verzoeken u het voorstel op de komende OLC vergadering te bespreken.

Mocht u nog vragen hebben, dan kunt u contact opnemen met Maurice van Keulen of Djoerd Hiemstra.

Met vriendelijke groeten,
Ida den Hamer
Secretariaat Leerstoel Databases
Zilverling 2015
Postbus 217, 7500 AE Enschede
Tel.: +31 (0)53 489 3690
http://www.utwente.nl/ewi/db/

Data Science (DS)

Course concept

Maurice van Keulen

1 Introduction

As a result of the EWI reorganization, the teaching load of our teachers is too high. Simply abolishing some courses is not desirable. We should strive for maintaining a variety of topics in our master curricula and that important topics are not missing. The way forward is, in my opinion, reforming our teaching forms in such a way that it better exploits the available resources: (experienced) teachers, PhD students and postdocs, and student assistants.

This course proposal is meant to reduce teaching load by merging two courses into one and by switching over to a different teaching form that is more easily scalable. In this way, we need not compromise too much on the variety or depth in topics we offer. Moreover, it is a form in which additions and changes to its content are more efficiently made. The teaching form is primarily based on problem-based learning and gets its inspiration from the first TOM-module 'Parels der Informatica' (PdI).

The two courses are Data Warehousing and Data Mining (part of Master BIT and bachelor TI) and XML & Databases 1, but as section 7 Variations and outlook describes, the merger of the two course is only an initial step in getting this course of the ground for 2014/2015. We intend to expand on this, hence also the broader title of the course. The course would be well-placed in Q2 or Q3 (the current place of DWDM and XMLDB1, respectively).

This memo gives an overview of the course concept. A more detailed version will be submitted to the OLC's of BIT and CS later.

2 Concept

To avoid misunderstanding, I would like to first explain that I make a clear distinction between *problem*-based learning and *project*-driven learning. In problem-based learning a well-described *closed* problem or assignment is leading. Any theory or course material is selected to aid the resolution of the problem. The path or paths towards resolution form a part of the design of the course, hence the teacher is in full control of what the student learns. In project-driven learning, an *open* project description is leading. Since the student determines the objectives of the project, the teacher can challenge and motivate the student to learn a lot, but has less control over what specifically it is that is learnt. Any related study material is usually discovered and studied within the project itself. Both forms are useful, but it is an important design decision which of the two one chooses.

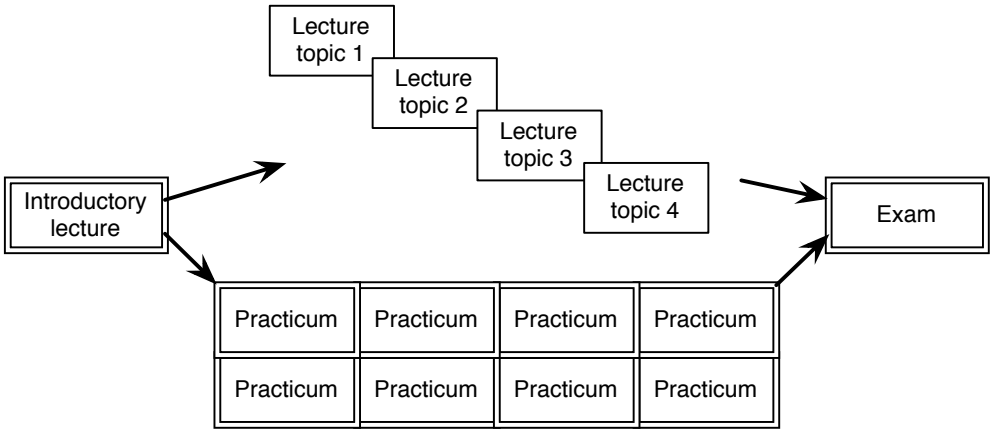

Figure 1: Sketch of the contact hours. Double borders indicate lectures shared by all topics

In short, DS is problem-based. It offers four topics (initially) and for each a well-described assignment is leading (see section 3). We strive for the same rigor in the description of the assignment as in PdI. Associated with each topic is 1 lecture focusing on only that part of the theory behind the topic that is not readily learnt by doing the assignment. Students are expected to choose two topics out of the four; each is "worth" 2.5 EC. Supervision for the assignments of all topics is combined: 1 teacher (the coordinator) assisted by one PhD student and in one lecture room. The coordinator also provides 1 combined introductory lecture. The written exam provides for a combined assessment: each topic exam contains 1.5hr worth of questions on theory and understanding. The assignments are supposed to contain self-assessment points, i.e., intermediary results where the students can check for themselves if what they have done so far is correct. Moreover, the assignments contain optional bonus assignments for which bonus points can be earned.

Besides the coordinator, several experienced teachers with different expertise are expected to be involved in DS: the responsible person for a topic (design of the topic assignment, 1 topic lecture, exam questions and grading). For the rest, i.e., the organization of the course and the practicum sessions are performed by 1 coordinator assisted by a PhD student. This keeps the load on the topic teachers low. Also additional topics can be added easily; the whole organization is already in place, the effort in the running of the course doesn't increase significantly, and most of the effort of the experienced topic teacher is a one-time effort. We start with four topics, but the selection of topics is expected to grow and change yearly. For deepening the knowledge on any of the topics, the REDI-course is available, which is a *project-driven* course where the student carry out a small research project.

3 Topics and learning goals

We have chosen the four topics below as the initial set of selected assignments.

3.1 XML Databases (Topic teacher: Hiemstra)

This topic contains about half of the learning goals of the course XML & Databases 1 (van Keulen / Hiemstra). The topic is not cutting-edge research anymore. This topic will teach the students how to publish XML-data from relational databases (SQL/XML), how to query and update XML data (XQuery), how native and relational-based XML databases achieve scalability, and how to do information retrieval with XML data. The assignment will consist of transforming a given data set in various ways and querying it.

3.2 Semantic Web (Topic teacher: van Keulen)

This topic contains the Semantic Web-part of the earlier abolished HMI-course Knowledge Representation (van der Vet / van Keulen). It will teach the students what Linked Open Data is, parts of the Semantic Web standard (RDF, RDFS), the description logic theory behind it and its limitations, and querying RDF data with SPARQL. The assignment will consist of an ontology modeling part and a querying part.

3.3 Data Warehousing and Data Visualization (Topic teacher: Amrit)

This topic contains the data warehousing, data visualization and multidimensional modeling parts of the course Data warehousing and Data mining. It will teach the students about data warehousing architecture, multidimensional modeling, and data visualization. The assignment will consist of developing a data visualization for a given case and data by means of modeling a star schema for the case, transforming the data set with ETL (we will use the tool Kettle) to provide input for a data visualization environment (Qlikview).

3.4 Data Mining (Topic teacher: Poel)

This topic contains the data mining part of the course Data warehousing and Data Mining. It will teach the students about classification, clustering, and association rule mining. The assignment will consist of performing these forms of data mining for given data sets using the tool Weka.

4 Assessment

The assessment of each topic is based on two components:

1. A written topic exam, and
2. The completion of the assignment.

The final grade for the course is simply the average. We intend to explicitly register the topic grades as 'deelcijfers' in Osiris.[1]

For good students to distinguish themselves, DS has a bonus system. A topic grade can be increased by doing the bonus part(s) of the assignment. If they are completed satisfactorily, up to 2 bonus points can be earned (while maintaining a maximum of 10 for the topic grade). The assignment is otherwise evaluated with a complete/incomplete only. Assignments will be signed off during the practicum itself by means of demonstration.

It is known, however, that students tend to ask for deadline extensions. In this course there will be no such deadline extensions. If the assignment is incomplete at the end of the quartile, the teacher will evaluate the incomplete result and decides either on a complete failure (for a severely incomplete result) or on a subtraction of up to 1 point (for a minorly incomplete result).

We explicitly allow a student to do 3 topics in one quartile, or to take the course again in a later year while choosing 2 other topics.

5 Effort and contact hours

Currently DWDM has 8 lectures given by 3 teachers and 1 PhD student with several overlaps, it has 2 homework assignments and a written exam. In 2013/2014, the teachers were present at 3, 4 and 5 of the lectures, the PhD student at 1 lecture, hence in total an effort of 12 * 2 / 2*2 = 24/4[2] contact hours.

[1] It is not known yet how Osiris handles 'deelcijfers' that are not all compulsory.

[2] "x/y" means x contact hours for the teachers and y for the PhD students

Currently XML & Databases has 8 lectures by 2 teachers with no overlap, it has 2 homework assignments and a written exam, hence in total an effort of 8 * 2 = 16/0 contact hours. Both combined is 40 contact hours (excluding the evaluation of the 4 homework assignments and the exams)

The new course concept asks for 8 practicum sessions for coordinator and PhD student and 1 introductory lecture for the coordinator only (9 * 2 / 8 * 2 = 18/16 contact hours), and 4 lectures for the 4 topic teachers (8/0 contact hours). This is in total 26/16 hours, which is a reduction to 60% on the contact hours of the teachers. Furthermore, there is no evaluation of homework assignments needed outside of these hours, because the assignments are supposed to be signed off during the practicum itself.

Adding an extra topic in this course concept requires a one-time effort of designing an assignment, preparing one lecture and designing an assessment approach. The running of the extra topic only requires 1 lecture = 2 contact hours for the topic teacher and making and grading the exam questions. The rest of the running of the topic is already in place and is combined with the running of the other topics. In this way, in subsequent course years, additional topics can be easily added and less-desirable ones eliminated.

6 Transition arrangements

Students who have already followed the courses DWDM or XMLDB1 but who have not passed it, will be offered exam opportunities as required by the OER. If the number of participants for such an exam opportunity is 1 or 2, we will organize an oral exam instead.
Students with one of both courses in their study plan, but who have not followed it, can simply replace it by DS while they choose the appropriate topics (3.3 and 3.4 for DWDM; and 3.1 with one of the others for XMLDB1).
Students with both courses in their study plan and have not followed them before, can replace them by DS and another course, or do DS twice and in this way complete all four topics.

7 Variations and outlook

As mentioned before, it is expected that in subsequent course years, more topics are expected to be developed, typically topics close to the research of the various chairs. If the course grows in this way, more students are expected to participate and more supervision can be needed; this is, however, simply a matter of adding one more PhD student.
One may consider to expand the course over two quartiles, i.e., add 8 more practicum sessions and distribute the topics over both quartiles. For 2013/2014 this is not deemed necessary, but when the number of topics grows and topics arise that build on another topic, this may be a logical way to go. It also makes it easier for students to do more than 2 topics.
The question can be posed “Why not two courses with each two assignments?” This would in my opinion unnecessarily reduce the flexibility and scalability.