13

Determining the Effect of TLAs on Student Engagement, Activity, and Understanding in a Repeated Teaching Setting

Toine Bogers

Det Informationsvidenskabelige Akademi, Faculty of Humanities, University of Copenhagen

Introduction

Search engines such as Google, Yahoo! and Bing, have evolved into complicated systems with many interrelated components. However, the basics behind their inner workings are relatively straightforward. The challenge in teaching students how search engines work is to break down the perceived complexity of search engines and make the students see through the complicated mathematical models down to the main conceptual steps.

How search engines work is one of the topics in the course *Digitale videnssystemer* (DV) that I have co-taught at *Det Informationsvidenskabelige Akademi* (IVA) for the past three years. DV is organized around three different topics that are all semi-related: (1) indexing and categorization, (2) bibliometrics, and (3) search engines. I am responsible for three lectures on search engines and it is the final lecture in this series that serves as the context for my final adjunktpægagogikum project. The topic of the final DV lecture is specialized search engines, such as recommender systems, question-answering systems, and expert search engines. Teaching this specific topic of expert search is what I wish to focus on in my adjunktpægagogikum project. Expert search engines are search engines that allow users to search for people that are knowledgeable about a particular topic (as opposed to, e.g., Google, which allows you to search for Web pages).

More specifically, I want to investigate the value of different teaching/learning activities (TLAs) for teaching my students about expert search engines, and find out which TLA is most effective in terms of engaging the

students, increasing their activity level in-class, as well as their understanding of the topic (in terms of reflection). This leads to my problem statement (PS) for this project:

PS What is the effect of different TLAs on student learning of the workings of an expert search engine?

Student learning as such is a broad concept and in this report I will attempt to address it using three research questions (RQs). One aspect of student learning that I wish to address is the concept of *student engagement*:

RQ 1 What is the effect of different TLAs on student engagement?

Engagement does not have a single, static definition, but as an appropriate working definition for this project report, I will use Chapman (2003), who defines it as the students' cognitive investment in, active participation in, and emotional commitment to their learning (Chapman 2003).

In-class activity levels are a specific part of Chapman's definition of engagement. However, because *student activity* is so often seen as being conducive to student learning as well as being one of the most directly observable outcomes for a teacher, I wish to focus on it specifically in my second research question:

RQ 2 What is the effect of different TLAs on student activity levels?

When answering this question, I want to focus on not just the teacher's perception of activity, but also on the student's perception of their own activity level as well as that of their fellow students to get a more complete picture of the effect of the TLA.

Finally, the goal of any TLA is to make students obtain knowledge and understanding of the topic(s) being taught. For this reason, the last aspect is wish to focus on through one of the research questions is *student understanding*:

RQ 3 What is the effect of different TLAs on understanding (in the form of reflection)?

I will attempt to answer these research questions by taking advantage of the structure of DV, which is divided into five different groups of 25-

30 students. DV has *sequential co-teaching*, which means that a different teacher comes in to cover each of the three topics. These five groups are taught by these three teachers in succession: teacher 1 has all five groups the first four weeks, teacher 2 the second three weeks, and I have all five groups at the end for three more weeks. This means that *within* each teacher's respective teaching period, repeat teaching takes place where each week the lecture content is repeated four times so that all five groups are exposed to the same lecture content. However, this also allows for different groups to be exposed to different TLAs, as long as the lecture content stays the same. I want to take advantage of this setting to gauge the effect of different TLAs on student engagement, activity, and understanding.

The remainder of this project report is organized as follows. The next section presents a brief overview of work related to the topics of engagement, activity and understanding and how different TLAs affect these. The section afterward presents the methodology of the aforementioned semi-controlled experiment with TLAs in more detail. That section is followed by a presentation of the results of this experiment with regard to my research questions. I conclude by discussing my findings and their implications for my future teaching.

Related work

Measuring the effect of different types of TLAs on students is not entirely new. For example, Andersen (2010) performed a similar controlled experiment when he compared a more traditional combination of lecturing and exercise classes with student-centered teaching, where brief lecture segments were intertwined with brief exercise segments (Andersen 2010). However, his experimental structure was a within-group design where the same students were exposed to two different types of lectures and quizzed at the end. This allowed him to compare the effect of these lecture types with the same group of students and directly measure their progress. My study is different in that it focuses on a specific TLA element in a larger lecture and that it takes place in a between-group design where I can concurrently measure the effect of different TLAs, ruling out any possible learning effects, as opposed to the more common sequential nature of Andersen's study.

Measuring student engagement with courses and individual lectures is not the easiest of endeavors, despite being deemed extremely important by many for effective student learning. One reason for this is that our under-

standing of what it means for a student to be engaged in teaching and learning has evolved over time (Chapman 2003, Nystrand & Gamoran 1991, K. A. Smith & Johnson 2005, Zepke & Leach 2010, Biggs & Tang 2011). Chapman (2003) presents a clear and concise overview of the evolution of engagement (Chapman 2003). In this report, I elected to stick to her own working definition of engagement as the students' cognitive investment in, active participation in and emotional commitment to their learning. Zepke & Leach (2010) offer a list of ten proposals for increasing student engagement. Two of these proposals were selected in particular for this report, because of their practical nature and manageable scope: (1) "enhancing students' self-belief", and (2) "enabling students to work autonomously, enjoy learning relationships with others and feel they are competent to achieve their own objectives" (Zepke & Leach 2010, p. 169). The two treatment TLAs that are described in more detail in follwing section aim to incorporate these proposals.

Methodology

The unique structure of DV allowed me to attempt to answer my research questions using a controlled experiment. While the lecture content of the final lecture always stayed the same, I could expose the five groups of students to three different types of TLAs when covering the topic of expert search. I measured student engagement and activity levels through direct observation as well as a survey administered at the end of the lecture. Understanding of the material was tested by including a reflection question in the survey.

The following sections describes the experimental setup in more detail, as well as the survey development and deployment. The sections after offer some more background information on the students taking DV and on the lecture in question.

Experimental setup

With the repeat-teaching structure of DV, the five groups of students could cover the same lecture content, but be exposed to different TLAs in a between-group design. I compared the effect of three different TLAs (or treatments), each corresponding to familiar paradigms from the pedagogical literature:

- **Traditional lecturing** In the traditional lecturing format I explained to the students how an expert search engine works, supported by slides and Q&A along the way. This corresponds to a *transmission-based learning* approach (Biggs & Tang 2011). Student group 1 was be exposed to this condition and served as my control group. To make sure their lecture lasted as long as for the other four groups, they were exposed to a longer group exercise on question-answering systems (QA), one of the other topics of the last lecture, to compensate for the missing expert search exercise.

- **Guided exercise** In the guided exercise the workings of an expert search engine were broken up into four different steps. For some example expertise areas, students were asked to go through each of the steps to produce a ranking of experts, which was then compared to the correct ranking. This way the students (hopefully) learned that a complex system can be broken down into simpler components and steps. This corresponds to a *problem-based learning* approach (Zepke & Leach 2010, Biggs & Tang 2011). Student groups 2 and 4 were exposed to this condition and served as treatment 1 The guided exercises was the default situation for all student groups in previous years.

- **Open discussion** In this format, I asked the students themselves to use their knowledge of how a search engine works—which was discussed in the two previous weeks—to conceptually design how an expert search engine could work and what steps would be involved in this. This corresponds to a *cooperative learning* approach (Zepke & Leach 2010, Biggs & Tang 2011). Student groups 3 and 5 were exposed to this condition and served as treatment 2.

All students received some introduction to what expert search is and why it is useful. The guided exercise and group discussion then took place in the treatment groups *before* the students were told how an expert search engine works. All students then received the same lecture segment on the two most important models of expert search *after* the exercises concluded.

Unfortunately, due to poor planning of overlapping deadlines of the different 3rd-semester courses, attendance was low near the end of the week due to an imminent term paper deadline in another course. While several students took part in one of the earlier lectures, the final two lectures had to be cancelled due to low attendance. This meant that for each of the three

TLAs I only had one student group instead of the planned two groups for treatments 1 and 2.

Evaluation

To measure the effect the three different TLAs had on student engagement, activity, and understanding, I used a combination of direct observation and a survey administered at the end of the lecture. While other methods would have also been useful to uncover the effect of the different TLAs, such as interviews or focus groups, I chose a combination of a survey and direct observation, because of the time pressure both the students as well as the teacher were under at the end of the semester.

Direct observation

While engagement and understanding are harder to evaluate through direct observation, it is possible to gauge student activity in-class in this manner. Student activity was observed by the teacher during all lectures as well as by a colleague during the group discussion lecture.

Survey

The goal of the (anonymous) survey was to measure student engagement, activity, and understanding, and as such consisted of three parts, each one corresponding to one of these aspects. For the engagement and activity parts, I asked both general questions and questions specific to the particular treatment variant (guided exercise or group discussion). All questions were in the form of statement the students were asked to (dis)agree with using a five-point Likert-scale. In addition to the TLA-specific questions, I also included questions in the engagement part about all the TLA types employed during the lecture to ensure that I would be able to compare all types of TLAs and not just the two treatment TLAs.

The students in group 1, who were exposed to the traditional-lecturing approach, received the same questions as the other four groups, even though they were not exposed to the special exercises on expert search. Instead, the four topic-specific questions were replaced with questions focusing on QA instead of expert search.

Because of the likely difficulties in stating whether a particular TLA contributed to a student's own understanding of the lecture content, I asked

no comprehension or understanding questions specific to a TLA, just general question. I tested their understanding by including a reflection question at the end of the survey. I then graded the students' answers on their quality on a three-point scale: good, medium, and bad.

The survey was deployed at the end of the lecture. The survey was only made available in English to avoid nonnative translation errors skewing the results (although English comprehension could of course have gotten in the way of understanding. The survey was made available to the students in both a paper and an electronic version. Students were free to pick their version of choice, but all students were asked to complete the survey before leaving the lecture. Appendix A contains the paper version of the survey.

Student characteristics

All of the ca. 150 third-year students have to take DV and at the start of the semester they were therefore divided up into five classes of around 30 students each. The majority of the students were in their second year of studying at IVA. The type of students that choose to study Library & Information Science (LIS) typically do not have a high aptitude for mathematics. While this in and of itself is not a problem, it is relevant when teaching a subject like search engines. The algorithms that make up a search engine are typically described using mathematical formulas, such as the calculate of term weights that tell the computer which terms are most important for a document. Experience has taught me that I need explain these formulas carefully and step-by-step (if it is necessary at all). I also try to reassure the students that are intimidated by the (arguably little and simple) math in the required reading that, while I will explain the math in as simple terms as possible, it is the principles behind these algorithms that are most important.

The teaching language for me (and for one of the other DV teachers) was English, although students were encouraged to ask questions in whatever language they were most comfortable with, Danish or English. This also applied to their term paper, which was the exam form of the course.

Lecture description

The topic of the final DV lecture was specialized search engines, which are search engines designed to perform a specific task or design to operate in

a specific domain. In this lecture I covered three examples of specialized search engine technology in the following order:

- *Expert search engines* (ES) are search engines that allow the user to search for people instead of documents. An expert search engine tries to automatically associated evidence of expertise (publications, social connections, project activity, etc) with experts and then rank them in order of perceived expertise. It was this topic that was the focus of my experiment with different TLAs.

- *Question-answering systems* (QA) are systems that directly attempt to answer questions asked by the user, such as "What is the capital of Paris?". Instead of the user having to transform such a question into a set of keywords, entering those into a search engine, going through the results list, and extracting the right answer(s) from these results, a QA system tries to automate these steps.

- *Recommender systems* (RS) are systems that attempt to recommend interesting items for future consumption based on past user preferences and/or purchases. A good example of a website that employes recommender systems is Amazon.com, which attempts to recommend other items to buy based on past purchases and purchases by others (e.g., "Customers who bought this, also bought ...").

13.1 Results

Figure 13.1 shows the answer distribution for all eleven questions relating to engagement, activity, and understanding over all 32 students combined that made up the three student groups. Bars in Figure 13.1 (and the other two figures) are color-coded by survey responses: green-colored bars represent (strong) agreement, while red-colored bars represent the (strong) disagreement. Deeper greens/reds represent stronger (dis)agreement. Median scores for each questions is represented by the 50% mark on the horizontal axis. In general, all students reacted very positively to the lecture: for five statements the median score was 5 and for the remaining six statements the median was 4.

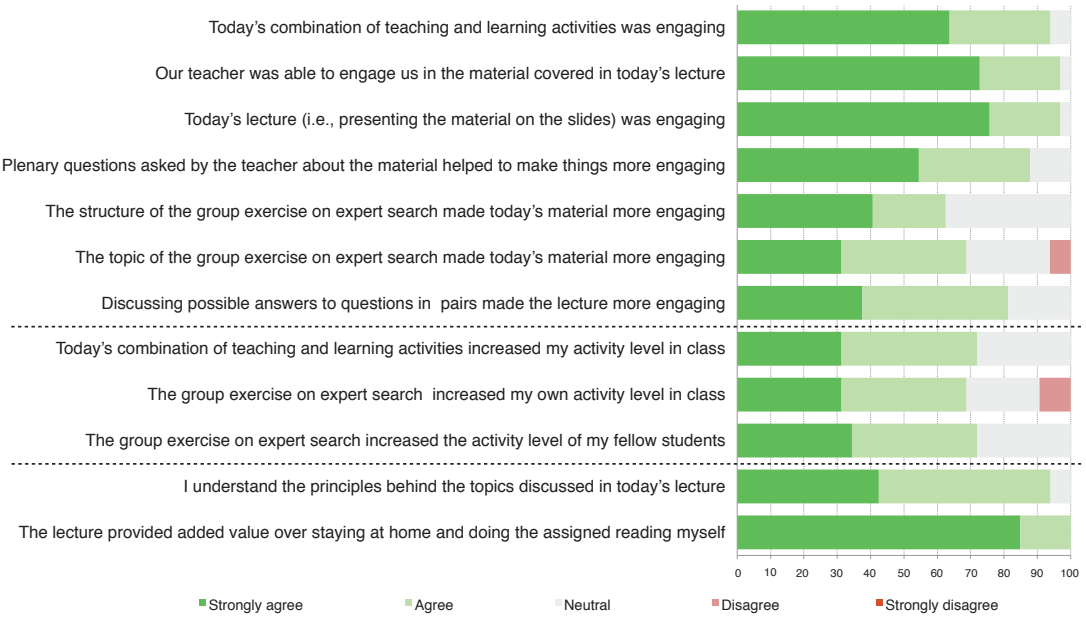

Fig. 13.1. Overview of the answer distribution for the twelve survey questions (N = 32)

Engagement

Overall, a majority of students stated that they felt (strongly) engaged by all types of TLAs. It is interesting to note, however, that the students actually were most positive about traditional lecturing as a TLA with around 75% of all students strongly agreeing with being engaged by this TLA. In contrast, the TLAs where students had to become more active by answering questions or discussing them with one or more of the fellow students scored slightly lower, although median scores were still all 4 or higher.

For the expert search-specific questions, I was interested in gauging their response to both the structure and the topic of the exercise to be able to separate the influence of these two factors on their feeling of engagement for the two treatment TLAs: the group-discussion variant vs. the guided-exercise variant. Figure 13.2 shows the answer distributions split by treatment type[1]: the group-discussion variant on the left vs. the guided-exercise

[1] It was not possible to compare the these two treatments to the control condition (traditional lecturing), because the lecturing-specific question was asked about the entire lecture and not just the expert search part. This means this was not directly comparable, weakening the setup of the controlled experiment. This should be addressed in future work).

variant on the right. From the distribution in Figure 13.2 it appears that students felt that the guided exercise engaged them more than the group exercise engaged the other group of students. The difference in median scores reflects this, with median scores of 4 for the guided exercise vs. median scores of 3 for the group discussion. However, when comparing the expert search-specific statements on engagement using a Chisquare test, it revealed no statistically significant relationship between treatment type and engagement as a result of the structure of the group exercise ($X^2(2, N = 22) = 2.011$, $p = 0.366$). Likewise, there was no statistically significant relationship between the type of treatment TLA and engagement as a result of the *topic* of the group exercise ($X^2(3, N = 22) = 2.377$, $p = 0.498$). Other comparisons between these two student groups or all three groups also revealed no statistically significant relationships between the groups.

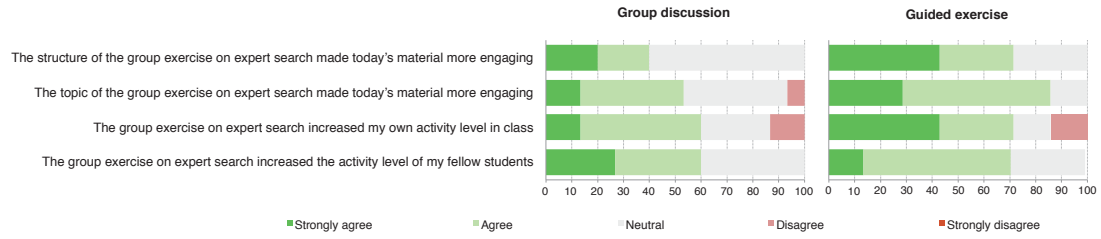

Fig. 13.2. Overview of the answer distribution for the four questions related to the non-lecture conditions 'group discussion' ($N = 15$) and 'guided exercise' ($N = 8$).

Activity

From the distribution visualized in Figure 13.1, it appears that students felt they were less active in class than they were engaged, which suggests that separating these two concepts was a good decision. However, when spliting their answers up by treatment TLA in Figure 13.2, there are some small differences in how students perceived their own activity vs. the activity of their fellow students. Students felt that while the guided exercise activated themselves more, the group discussion exercise seemed to activate their fellow students more. However, these differences are not statistically significant ($X^2(3, N = 22) = 2.538$, $p = 0.469$). Likewise, there was no statistically significant relationship between treatment type and activity of other students

as a result of the group exercise ($X^2(2, N = 22) = 1.155$, $p = 0.561$). Interestingly, both my observations and those of the external observer suggested that the group discussion exercise was actually the most successful TLA for activating the students. This suggests a possible disconnect in how teachers and students define 'activity' in class.

Comprehension

The bottom part of Figure 13.2 shows that students from all groups seemed very confident in having understood the principles behind the topics of the last lecture with a median score of 5. In addition, 84% of them strongly agreed with the statement that the lecture provided added value over staying home and doing the assigned reading for themselves.

When testing their actual understanding of expert search through a reflection question on how to incorporate the temporal dimension into an expert search engine, the results were not as overwhelmingly positive. Of the 32 students that answered the questions, only 31.3% submitted a good answer, while 50.0% submitted a medium-quality answer and 18.8% of the students would have failed, were this a real exam.

Figure 13.3 shows the answer distribution split by the three TLA types. Considering this split, it is perhaps surprising that, considering their feelings about their own engagement and activity, the group discussion students entered the highest number of good answers at 40%. And while the guided exercise-group judged their engagement and activity as the highest among all groups, they had the lowest number of good answers of all three groups. However, according to a Chi-square test there was no statistically significant relationship between the TLA type and the answer quality $X^2(4, N = 32) = 2.055$, $p = 0.726$).

Discussion & Conclusions

In this project report I have presented a small-scale pilot study on the effect of different types of TLAs—traditional lecturing, a guided exercise, and group discussion—on student engagement, activity, and understanding in a repeated teaching setting. When it comes to *student engagement*, students seemed to be feel more engaged by the guided exercise, where they were taken through a series of steps that signified how an expert search engine worked, than discussing how expert search engines worked in a group.

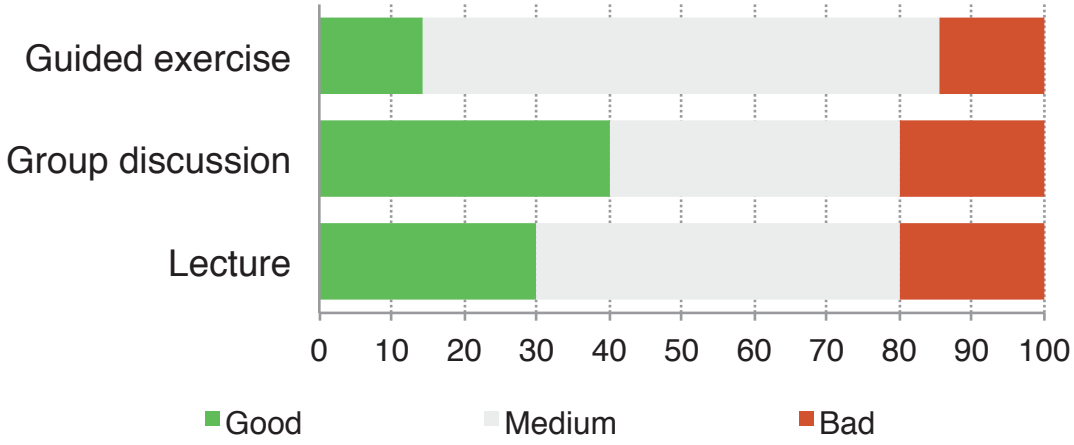

Fig. 13.3. Overview of the grade distribution for the three different conditions (N = 32).

However, overall students reported feeling engaged the most when being exposed to the traditional lecturing format. A possible explanation for this could be that students are simply more used to the traditional lecturing format and that discussing in groups makes some of them more uncomfortable and therefore less engaged. Another problem could be that students interpret the concept of engagement differently. While I purposefully did not define it beforehand, it does make it harder to compare the results.

Looking at the survey results, the effect of the type of TLA on *student activity* levels was not as marked as it was for engagement. An interesting finding in terms of activity was that while students felt that the guided exercise activated themselves more, the group discussion exercise seemed to activate their fellow students more. A possible explanation for this could be that with the guided exercise every single student has to contribute, as the students typically divided the different assigned expert areas amongst themselves. This meant that students were very aware of suddenly having to become active themselves, while the activity of other students became less visible to them. In contrast, group discussion makes it easier for a student to 'hide' in the group, while it makes the activity of their fellow students much more visible. From a teacher's perspective, however, direct observation suggested that the group discussion exercise resulted in the highest activity levels.

Evaluating the student's *understanding* of the material covered in the lecture showed a different picture than that of engagement and activity. In

addition to the group discussion format producing the most activity, students participating in this format also produced the best answers to that exercise, as they came up with exactly the answers I was hoping for as well as some original solutions I had not considered myself. This was also reflected in the answers to the comprehension question in the survey.

The results presented in this report should be taken as those of a pilot study: the number of students participating in this last lecture was too low to be able to draw statistically meaningful and representative conclusions. In addition, direct observation and surveys might not be the best methods to get at the students' attitudes towards engagement. Interviews or focus groups would make a good complement in future work in this area.

On a personal level, I plan to use the group discussion variant more extensively in the future, because of my positive experiences with the student's activity level and the quality of their answers during and after this type of TLA. However, this does not mean the other two TLAs are without merit. For instance, the guided exercise could easily be re-worked into an online quiz that would allow the students to test their understanding of the documentcentric expert finding model at their own pace, providing an additional check on their understanding of the material.

A Survey questions

Survey on teaching & learning activities in Digitale Videnssystemer

The goal of this survey is determine your attitudes towards specific different teaching and learning activities in today's teaching and how you experienced the lectures as a whole. Thank you in advance for participating!

Please rate the following statements on how you've experienced today's lecture for how much you agree or disagree with them.

		Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Engagement	Today's combination of teaching and learning activities was engaging	O	O	O	O	O
	Our teacher was able to engage us in the material covered in today's lecture	O	O	O	O	O
	Today's lecture (i.e., presenting the material on the slides) was engaging	O	O	O	O	O
	Plenary questions asked by the teacher about the material helped to make things more engaging	O	O	O	O	O
	The **structure** of the group exercise on **expert search** made today's material more engaging	O	O	O	O	O
	The **topic** of the group exercise on **expert search** made today's material more engaging	O	O	O	O	O
	Discussing possible answers to questions in pairs made the lecture more engaging	O	O	O	O	O
Activity level	Today's combination of teaching and learning activities increased my activity level in class	O	O	O	O	O
	The group exercise on **expert search** increased my own activity level in class	O	O	O	O	O
	The group exercise on **expert search** increased the activity level of my fellow students	O	O	O	O	O

(Please turn the page)

		Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Comprehension	I understand the principles behind the topics discussed in today's lecture	O	O	O	O	O
	The lecture provided added value over staying at home and doing the assigned reading myself	O	O	O	O	O

Finally, I would like to ask a question to test your understanding of one of today's lecture topics. This is of course not a part of your final grade. Nevertheless, please answer this question to the best of your ability and write your answer (either in Danish or English) in the box below.

Question:
Expertise on a particular topic X is very dependent on time: although a particular researcher might have been an expert on X twenty years ago, if (s)he has not done any research on X since then, her/his expertise is probably not that high anymore. Conceptually, how could we add this time dimension to an expert search engine to make it more realistic? Is there a particular stage or phase we could add this information?

Thank you for your participation!

All contributions to this volume can be found at:

http://www.ind.ku.dk/publikationer/up_projekter/2014-7/

The bibliography can be found at:

http://www.ind.ku.dk/publikationer/up_projekter/

kapitler/2014_vol7_nr1-2_bibliography.pdf/