Minutes to the INSITE Workshop at ECLT, Ca'Foscari University, 18-19 September 2012, Venice

Measuring, Quantifying and Modeling Processes of Scientific Knowledge Creation and Propagation in Online Societies

Vision: Changes in the world of learning. What will learning and teaching be in 20 years?

We view the process of knowledge creation as an ecosystem where learners, students, parents, teachers, university presidents, and ultimately society, all should profit from each other as knowledge is transmitted and/or created. New ways and possibilities of teaching and learning are at the edge of being successfully implemented, and might radically transform our view on traditional forms of learning and on the associated institutions. Traditional European-style universities are institutions that existed continuously for about 700 years and have helped to establish a Western Culture. What are the components that made them so successful? Next to the component of teaching practical (and unpractical) things to the next generation, universities fulfill a series of social purposes. They serve as a platform for meeting and establishing ties in the future generation of decision makers, and they serve as a ground for emancipation from ones background and family. To a certain extend they give the students the possibility to re-invent themselves, their role in life, and their goals at the age of 20. A university is composed of a set of building blocks that can possibly be re-invented to match and use present day information technology without losing the social and emancipatory components a university should have. In this context a few key questions arise:

- What is knowledge? Is it an absolute value or is it a social construct? What does the answer to this question mean for the use of IT methods in learning?
- In what ways can learning and teaching be assisted by IT components, like third-generation machine learning? What is an artificial teacher?
- Can a future digital university compensate for the limited social components? Can the social experience possibly be re-gained through new social media on the internet, or the use of geo-localization match-making?
- How can the knowledge-creation component of a university be made accessible and open to a wide public? How can it be used for talent shouting and optimal resource allocation? How can learning through research be made available to those not enrolled in elite institutions?
- Global vision: How much can a digital university be scaled up to?
- How does the narrative of education change over time: the 18th century narratives included enlightenment components, maybe even utopia-like ones. This changed more and more to an industrial, economical, utilitarian focus which is becoming more and more pronounced. What should the narratives of the 21st century be? Egalitarian?

How can science and technology contribute to these questions? What visions are necessary to confront the expected changes in learning, teaching and knowledge creation in optimal way? What does optimal mean? Present day IT will allow to *experiment* with different educational systems. Two major routes are imaginable: (1) Evolutionary approach: Leave the system as it is, and change one component at a time (mutation), and see what the consequences are (check fitness). See for example the initiative *The University of the People*, by Shai Rashef. There the University only provides rooms and meeting space, the teaching component is completely outsourced to online courses and sources. (2) Technocratic approach: design a new teaching environment and apply it globally. A successful example is the *coursera* program of Stanford University.

Workshop outcomes: The idea of the two-day meeting was to come up, discuss and harmonize several visionary ideas of the future of knowledge creation and learning. These issues could be used to trigger a discussion if these or similar ideas could prove useful for future EU funding in the Horizon 2020 framework or within initiatives like the FuturICT flagship. A first step toward a concrete research project was taken by identifying and clarifying potential workpackages in a project that would illustrate the potential of future personalized learning. The main topics discussed at the workshop included

1) The future of knowledge creation with the participation of citizens or non-traditional specialists
2) Ways of future knowledge creation and learning in the light of demographic and behavioral changes in the population. Examples from successful initiatives in mass education in India.
3) What is learning? What is it that potential on-line universities do not have? Is personalized (human-human) mentoring necessary?
4) New technologies as a pre-requisite of personalized learning: content based time-varying networks and emotional classifiers
5) Online Game university: vision for future online campuses and automated and personalized learning in online game environments
6) First steps in designing a joint research project in: interactive – personalized learning platform based on Wikipedia.

1) **The future of knowledge creation.** Knowledge creation in the traditional ways might reach a limit. The numbers of physics PhDs in the western world is stagnant since the 1970s. The number of available positions in research institutions and universities is not growing, and the limits of resources on research spending become obvious more and more. Even though there is still exponential increase of scientific output according to several measures, such as number of papers, etc., this is not sustainable given a non-growing number of traditional experts and researchers.
To keep and maintain an exponential growth of knowledge creation – which might prove helpful in confronting the big global problems (climate change, migration, over-population and social unrest, religious fundamentalism) new forms of knowledge creation might become unavoidable. In this context the first successes of Citizen Cyber Science, the involvement of thousands to million non-experts, have taken place.

2) **Ways of future knowledge creation.** There are various examples of Citizen Cyber Science (CCS), differing in the degree of individual involvement and the degree of coordination between the Citizen Scientists themselves and between them and the traditional science community. Famous examples for a weakly interacting CCS community is Galaxy-zoo, where individuals classify galaxies recorded from telescope data; a more interactive and coordination based activity is the game fold-it, where citizens try to find optimal configurations of protein sequences. Results of both examples are spectacular both in the number of participants as well as in the actual scientific output.

3) **What is learning?** Given new electronic tools, can it be understood what learning actually is? Given data on CCS initiatives, it is possible to see and reconstruct how individuals conceptualize and incorporate information they have recently acquired. It is possible to follow their paths to success in solving a problem, to see the rate of increase of efficiency in solving similar problems, etc. Given that in the near future vast amounts of data of this kind could become available and could be produced for that purpose, it is conceivable that it should be possible to make the process of learning an observable and quantifiable science. Understanding on that frontier, maybe in combination with third-generation machine learning technology, could become the basis for automated-personalized teaching.

4) **New technologies for personalized learning.**
As in personalized medicine it is expected that learning can be personalized through Bayesian models in machine learning. What is necessary for this approach is that the computer is able to gain information on the individual behavior of the student as she or he learns. This information in combination with large-scale databases containing the information of large collections of students' behavior and further individual information on students will make it possible for a Bayesian machine to guide and suggest next steps for the students (artificial teacher). This, in combination with existing technologies like video lectures, and courses like coursera, brings personalized learning within reach of the next decade. The combination of

students' behavior, interest, and progress rates, with geo-localization tools, could provide a tool to assist the possibility for physical meetings of students, who share similar interests, similar levels, similar backgrounds, etc. How can new technologies be used to educate the underprivileged – maybe illiterate? A tremendously successful project in India was presented, that uses cell phones for AIDS prevention and guiding pregnant women through pregnancy.

5) **A thought experiment: The Online Game University.**
Given the tremendous popularity of massive multiplayer online games and their growth rates, can one imagine a university designed as a computer game? Each player is represented by an avatar, who finds itself at the doors of a university, a miraculous place full of adventure games. By exploring the place the avatar gets drawn into quests which require him to learn (and memorize) a variety of things so that the quest can be accomplished. During the game the avatars meet others with whom they can explore, and learn together. Such a game university can be built in a modular way such that people can link their individually programmed educational quests to the existing game. In this way it is possible to learn about the learning of the avatars, i.e. a quantitative science of learning. The age range could be 4-99. The Game University must be an adaptive and learning system itself that adapts to needs of their users.

6) **A real project: Interactive and personalized learning on the basis of Wikipedia.**
As a possible concrete science, IT, and research project we suggest using Wikipedia as a model system for learning. Wikipedia is currently used by approximately 500 million users, basically as a source of learning. Even though Wikipedia is a starting point for learning and is the most dominant homepage in many countries it has a frustrating component attached to it. Since the content is mostly generated by experts, it is often not understandable by non-experts. Examples include the pages for elementary particles like the proton, or proteins connected to certain diseases. The fact that many people do not understand the majority of the content is frustrating and hinders learning and knowledge creation. The aim of the project would be to individually guide the users to Wikipedia pages that she or he is most likely to understand. The computer (a Bayesian machine) serves as a matchmaker of people and content they can understand. Even small progress in this direction will have a huge impact, due to the numbers of users included. The core of the project is an app that users willingly install on their computers, that allows the Bayesian machine to learn about the behavior of the user. The user also provides feedback of her learning experience to the machine. In this picture Wikipedia serves as a Knowledge base (network), users diffuse on this network. This usage is used to monitor learners, give guidance (teaching), suggest improvements in asking questions, identify common learning paths, connect people (serves as a match maker), and visualize – why do users want to use it? The computer would feed back what you as a user have explored, it would give back the *intellectual fingerprints* of those users that use the improved Wikipedia platform (‚know yourself'). Concrete components of this project would involve the following packages, most of which could be covered by the present workshop participants and their institutions.

- App to record learning paths: for individual use, shared use, Bayesian machine use, …
  The links from one page to other carries meaning and content. This must be quantified by new classification software and content-network technology.
- Link suggestion tool: the tool assists Wikipedia users to find pages of right expert level, if it exists. If it does not exist it issues a need for that page and suggests creating one.
- Tools for teachers: teachers are given learning paths of kids, so that they get a feeling for what the kids know – to facilitate a more personalized teaching in school.
- Who-Am-I-Pedia: an app maps your intellectual profile and gives it back to you.
- Ethics and privacy tool: allows you to specifically block your computer from learning about you.
- Research question: how does learning work? Learning paths from thousands of users will be analyzed for feedback mechanisms, popular foraging paths of learners, signs of self-organization, reinforcement, etc. Can a relation between the use of knowledge and decision making be established, i.e. can degrees of ‚rationality' be measured in different users?
- Research question: multiplex aspects of learning. How do social networks influence learning? Are alternative ways of selective pressure for the creation of Wiki pages possible, besides precision and content?
- Crowdsourcing tools for researchers: open API for scientists to facilitate Citizen Cyber Science
  Matchmaker tool for scientific projects: time, resources, advice, money.
- Global map of knowledge, personal map of knowledge: an app shows your intellectual fingerprints on the personal level, or those of your friends, communities, or the global one.

Benefits for a European engagement:

EC engagement in this technology based re-thinking of future learning and education could be of potential interest in a European context.

1) EC commitment could ensure that these new ways of learning and teaching remain *open* on several levels, and will not happen behind walls and doors of commercial companies. Offered to all citizens and information gained from them will be made open as well. Information about the learners will be fed back to the learners in a transparent way. This could be seen as a contemporary version of the ancient Greek maxime *‚know yourself'*.

3) EC engagement and successful outcomes could challenge the miserable state of the educational system in many countries of the EU, and hence could stimulate overdue reforms.

4) Enlightenment goals and values could be transported along with other European values. The EC could see its engagement as a way to propagate a new human right: right to know, and the right to access high quality knowledge.

5) Europe is once more falling behind: In 2012 Harvard University and MIT started the Projekt *edX*, which offers online seminars. More than 150.000 students from all over the world attended the first courses. Even more successful is the *Coursera* program of Stanford University, which has enrolled more than a million students from 196 countries. Successful European initiatives in this direction are missing.

6) It could help to lower the effective barrier for people past student age to participate in *life-long learning*.