

One experience in e-earning

Miljana Mladenović, ml.miljana@gmail.com
University of Belgrade, Faculty of Mathematics

During last few years, many universities have been offering, beside traditional education, different possibilities for studying by Internet. One can use many terms: on-line education, e-learning, distance-learning for the same thing. This type of education is marked by a special kind of relationship between professors and students based on communication via Internet services. The process was initiated by small companies and schools that were looking for ways of attracting more students. But, everything became more serious and professional when the world most famous universities, like MIT, Berkeley and Stanford, entered the process of on-line education. Today, besides regular on-line studies, there are also some public free courses. During the spring semester 2012,

the Stanford University has offered 8 courses to graduate students: Design and Analysis of Algorithms I, Probabilistic Graphical Models, Natural Language Processing, Game Theory, Software Engineering for SaaS, Machine Learning, Human-Computer Interface and Cryptography. All interested students had to register in early days of January by sending an e-mail. For all regular Stanford students these courses have the same status as traditional courses; for all other candidates that come from all over the world, it is an opportunity to acquire more knowledge in the field of their interest. Those who successfully complete the course will receive a statement of accomplishment.

University of Stanford publishes all of its courses in association with

Coursera, a non-profit organization specialized for on-line education. At every moment, on a page <http://www.class-central.com/> one can find information about all courses that are ongoing, upcoming or finished with information about lecturers, starting dates and course durations.

Intro	Course Name	Instructor(s)	Stream	Start Date	Length	Initiative
	Natural Language Processing	Chris Manning Dan Jurafsky	Computer Science	12th Mar. 2012	8 weeks	Coursera
	Game Theory	Matthew Jackson Yair Shoham	Computer Science	19th Mar. 2012	7 weeks	Coursera
	Probabilistic Graphical Models	Daphne Koller	Computer Science	19th Mar. 2012	10 weeks	Coursera
	Cryptography	Dan Boneh	Computer Science	12th Mar. 2012	6 weeks	Coursera
	Design and Analysis of Algorithms I	Tim Roughgarden	Computer Science	12th Mar. 2012	5 weeks	Coursera
	Model Thinking	Scott E Page	Complex Systems	20th Feb. 2012	12 weeks	Coursera
	CS 373: Programming a Robotic Car	Sebastian Thrun	Computer Science	20th Feb. 2012	7 weeks	Udacity
	CS 181: Building a Search Engine	David Evans Sebastian Thrun	Computer Science	20th Feb. 2012	7 weeks	Udacity
	EE 930x: Circuits and Electronics	Anant Agarwal Gerald Sussman	Electrical Engineering	9th Mar. 2012	14 weeks	MITx

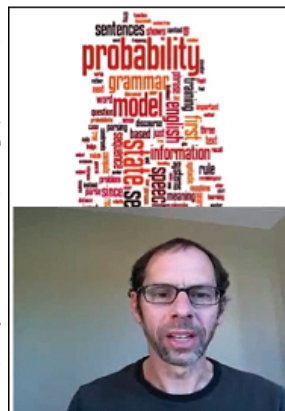
My domain of interest is Natural Language Processing - NLP. So, I didn't wait, but sent at once my e-mail registration message for Natural Language Processing course for which lecturers were prominent professors Chris Manning and Daniel Jurafsky. The planned starting date for this course was January 23th, but since more than 45 thousand candidates applied for this course, organizers had to improve the whole system in order to accept such a huge number of students from all over the world. Finally, on March 12th, lectures started. University of Stanford considered this an important event, so it published the news at its site <http://www.stanforddaily.com/2012/03/06/stanford-partners-with-coursera-to-offer-more-online-courses/>. Meanwhile, students received information about obligatory literature. The recommended books were: „Speech and Language Processing“ by Daniel Jurafsky, „Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition“, Daniel Jurafsky and James H. Martin and „An Introduction to Information Retrieval“ by Chris Manning and associates. In addition, some links to programming tools, web resources and applications were specified. Lecturers drew students' attention to

Java and Python as the only programming languages that could be used in problem resolving. Syllabus and plans were presented, due dates for homework and problems' solutions were specified, the evaluation system was determined, students received user names and passwords as well as instructions how to send solutions to the problems. The course started at March 12th.

Following the syllabus, in a first week of NLP course there were several interesting lessons: using regular expressions in practical NLP, word tokenization, word normalization, sentence splitting and stemming. The notion of a minimum edit distance was introduced as well as algorithms to compute it. A term „weight“ was defined and algorithms that use them were tested. At the end of the week, the use of these algorithms in Computational Biology was presented. All materials were divided in 10 lessons followed by 10 video lectures, as well as corresponding ppt and .pdf files for those who prefer reading to listening to lectures. Personally, I prefer lectures „by word of mouth“, especially when a professor sketches on a blackboard. It seems like being in a „real school“, except that there are no colleagues around me, chair crunch can't be heard and one cannot hear a bell that announces the end of one class.

Two types of evaluations follow each lecture: Problem Sets and Programming Assignments. The first one is a well known ABC-test. A student has to answer to 4 or 5 questions. For some of the questions only a box has to be checked, for others some calculation has to be done on a local computer and the result has to be typed in a text box. Finally, for some questions, student has to write a program, executes it and write a result in an appropriate box. A due date for all students to send their homework is seven days from a moment when lessons were published. Another 7 days is a grace period. After that time it is practically useless to send them because penalty points amount to 50% of earned points. Besides that,

there are new Problem Sets and Programming Assignments for each week. A student unsatisfied with the Problem Set results, can repeat this test up to 5 times by receiving new set of questions. The best score is accepted as a final. Problem Sets are some kind of repetition of given lectures. But, the big challenge is Programming Assignments. They are computer programming homework. All of them are designed as requests to solve a given problem in the optimal way. Students have to write programming code to satisfy some distinctive parameters. In my opinion, this type of assignments differs very much from those usually given at Serbian universities. At our faculties students often have to solve some real problems, while here, the problems to be solved are much more tricky, and sometimes it seemed to me that I was walking through a labyrinth. Here is an example: our first assignment had a mystical name SpamLord. Students had to parse a bunch of web personal pages of Stanford University professors and associates and try to find all their masked e-mail addresses and phone numbers. An email address can be masked in many different ways. For example, as: `pera at matf dot bg dot ac dot rs` or `pera (at) matf;big;ac;rs` or `pera where matf dt bg DOT ac;rs` etc. Also, we had to think about extraction of phone numbers. When we caught all e-mail and phone number candidates, we had to transform them in appropriate and useful known shapes: `address@domain` or `xxx-xxx-xxxx` for phone numbers. Lecturers evaluated our programming codes by calculating false positive and false negative hits, so number of earned points grew up as that two parameters fell down. As the process is iterative and not a trivial one,



one must develop a strategy and make compromises in order to optimize a solution, it is important that a student remains optimistic and opened for further research. I must admit, that I was a bit of a sceptic at the beginning of the course, especially when the programming assignment was given after lessons in the second week. Four different algorithms for so-called language models had to be implemented. That is, probabilities of word frequencies had to be estimated by using a training model. After creating a language model, it had to be applied on both development and test models. Evaluation parameter was based on obtaining the accuracy greater than it had been prescribed. I was left in the depths of despair by accuracy zeroes at the start of problem solving. But, one of the best things of on-line education is Students discussion forum where you can meet many other students from different countries and cities that have the same or similar problems. So your problems become smaller and joy becomes bigger when you successfully finish your task and share your success with all of them. On the forum you can advise somebody who “overslept“

Minimum Edit Distance

Computing Minimum Edit Distance

the maths classes that $\log(a/b)=\log(a)-\log(b)$, and somebody can advise you that you should include the infamous sentence tags `<s>` and `</s>` in your calculations.

So, the NLP course goes on. There are no big differences compared to regular courses. But, there are no student cafeterias, fifteen minutes breaks and a bunch of interesting people around you. Everything in life has some good and some bad points. Our duty is to maximize the good ones and to minimize the bad ones. Fortunately, that can still be done without computer programming.