Aalto University

Aaltodoc
OPEN ACCESS

Author(s): Lahti, Lauri

Title: Experimental evaluation of learning performance for exploring the shortest paths in hyperlink network of Wikipedia

Year: 2014

Version: Post print

Experimental evaluation of learning performance for exploring the shortest paths in hyperlink network of Wikipedia

Lauri Lahti

Department of Computer Science and Engineering
Aalto University School of Science, Finland

Abstract: In a 9-hour experiment we evaluated learning performance based on exploring the shortest paths in hyperlink network of Wikipedia online encyclopedia. Relying on network of 35688 unique hyperlinks in three separate learning sessions of 20 minutes students read series of 62 sentences built by using 22 unique hyperlinks that form the eleven shortest paths and answered pre-test and post-test multiple-choice questionnaires about recall of sentences (tests 1-6). For experiment group (n=24) 62 sentences were chained in such an ordering that corresponds to traversing cumulatively a series of associative trails leading from concept Tourism in Malta to concept Euro coins of Malta along alternative parallel shortest paths in hyperlink network of Wikipedia category Malta. For control group (n=10) same sentences had randomized ordering. For both unique hyperlinks and consecutive pairs of hyperlinks experiment group reached higher degrees of recall than control group in tests 2-5 and the effect size in favor of experiment group was over 0.18 for test 2 and over 0.40 for tests 3-4. We do not know any previous work verifying learning performance like in our approach.

Introduction

We carried out an experimental evaluation of learning performance for exploring the shortest paths in hyperlink network of Wikipedia. We aimed to carry out experiments with a sufficiently compact educational topic not previously well known among the students and thus with a cultural-historical perspective we selected the topic to be European state of Malta. We extracted in June 2014 a subsection of hyperlink network of English language version of Wikipedia about state of Malta so that we first extracted all 3278 Wikipedia articles belonging to Wikipedia category Malta (http://en.wikipedia.org/wiki/Category:Malta). These 3278 articles had 226329 departing hyperlinks and of them 185610 were unique. Among these 185610 unique hyperlinks 20757 had an end concept belonging to group of 3278 Wikipedia articles belonging to Wikipedia category Malta. After allowing all 20757 unique hyperlinks to be traversed in both actual linking direction and in opposite direction and after eliminating those hyperlinks that are connected to article "Index of Malta related articles" we finally got 35688 unique hyperlinks. We do not know any earlier work verifying learning performance like we have now done. Wikipedia has properties of scale-free small-world networks (Masucci et al. 2011) and we hope that our findings can motivate further research about learning processes relying on exploring the shortest paths in network that has properties of scale-free and small-world networks. A more extensive analysis of the experiment and detailed listings of used data are available in publication (Lahti (to appear)).

Previous research

In adaptive educational systems learning performance curves relying on power law of practice or alternatively exponential curves can be used successfully in formative studies (Martin et al. 2011). An experiment with vocabulary learning in a self-guided web-based language learning environment showed that constructed responses items had greater effect than the multiple-choice items on posttests about recall and recognition of the students (Chen & Chen 2009). However, there are also claims that constructed-response questions are equal to multiple choice questions that allow multiple responses and use scoring rule counting only correct responses (Kastner & Stangl 2011).

It has been argued ((Thalmann 2014) referring to (Akbulut & Cardak 2012)) that that continuum of adaptation needs of adaptive systems and how to match them to suitable learning materials has gained limited attention in research and main focus has been on learning styles rather than other factors. Liao et al. (2012) argue that in research of game-based learning many empirical studies have focused on learning outcomes rather than learning process. To provide guidance for a learner with an adaptive intelligent tutoring system typically requires parametrization of learning process that relies on some theory like item response theory approximating learner's answers with a probabilistic function (Mohamed et al. 2012). In respect to student interacting with intelligent tutoring systems and self-regulated learning, Bouchet et al.

(2013) argue that while there is an excessive amount of research of student models, most of it has not emphasized complex cognitive, metacognitive, motivational and emotional processes with non-linear hypermedia learning materials.

Using a semantic network model based on Wikipedia, Thompson and Kello (2013) showed with recall experiment that semantic memory processes can be modeled as searches over scale-free networks and inter-retrieval interval became progressively greater as minimum path length increased between nodes of semantic network to be recalled. Gureckis and Markant (2012) offer a review how process of self-directed learning can be motivated with both cognitive and computational perspectives thus contrasting human memory processes and machine learning methodology. Relatedness and dependency of sentences in text can be computationally represented by spreading activation in a network of sentences and this approach has been used successfully to predict the student's self-explanations concerning bridging and elaboration based on text cohesion (Bellissens et al. 2010). Tintarev (2009) lists five features in respect to evaluation of recommender systems including accuracy, coverage, learning rate, novelty and serendipity, and confidence that is often in relation to strength.

It has been difficult to convert an existing system to manage a different learning topic, and thus self-improving intelligent tutoring systems have been suggested (Soh & Blank 2008). It has been argued (Robberecht 2007) that computer-based learning materials seem to force learners to single-mode pedagogy with linear or sequential design although they should promote nonlinear, interactive practices with context-sensitive and active learning elements accommodating diverse learning levels and styles. In adaptive hypermedia systems adaptation techniques have been categorized into adaptive presentation and adaptive navigation support so that adaptive navigation support further consists of direct guidance, adaptive link sorting, adaptive link hiding, adaptive link annotation, adaptive link generation and map adaptation (Brusilovsky 1996).

Persons having cognitive style of field dependence are claimed to seek external referents for knowledge and be better learning human-related content whereas persons having cognitive style of field independence are claimed to develop own internal referents for knowledge and be better learning impersonal abstract content (Chen 2002). It appears that Western style cultures are more field independent than Eastern style cultures and that along growing from childhood to early adulthood people experience decrease in field dependence but it may later in life increase again (Thomson et al. 2014). In respect to learning achievement and learning time with computer-based instruction program, field independent students performed better with non-linear presentation and field dependent students better with linear presentation, and it has been suggested that hypermedia learning programs can help field independent learners by offering multiple routes, free choice and visual control, and can help field dependent learners by offering guided routes, labelling the role of current position along path and separate directions to required information (Chen 2002).

In our previous work (Lahti 2014) we experimentally measured recall of selected hyperlinked concepts and shown hyperlinked concepts in hyperlink network (which can be considered to support field independent learners), and in our current work we experimentally measured recall of shown hyperlinks forming the shortest paths in hyperlink network (which can be considered to support field dependent learners).

Method

To carry out a sufficiently compact educational exploration task in just described hyperlink network of 35688 unique hyperlinks we decided to observe how students explore the shortest connecting paths leading from concept Tourism in Malta to concept Maltese euro coins. We made an assumption that an average student knows that Malta is a state that can be visited as a tourist and that this knowledge can be used as a useful starting point of exploration path for adopting more knowledge about Malta, and we chose Maltese euro coins as the ending point of exploration path since we considered that euro coins are some concrete objects related to Malta that can be encountered in everyday life of students and thus can make learning topic more personally touching and meaningful (students participating in our experiment live in euro zone and thus can sometimes notice also Maltese euro coin versions among locally circulating euro coins). To generate the shortest paths in hyperlink network showing them in decreasing order of length of the path and also showing all alternative parallel paths having equal length we used Yen's algorithm to compute top k shortest loopless paths with sufficiently high values of k (Yen 1971).

It turned out that in our hyperlink network the shortest path is Tourism in Malta → Malta → Maltese euro coins but we wanted to create an exploration task that covers the chosen learning topic Malta

more broadly. Therefore we decided to take into consideration only shortest paths having the length of three hyperlinks, thus getting 29 alternative shortest paths, and additionally excluding all shortest paths that traverse concept Malta we finally had *11 alternative shortest paths* containing 13 concepts[1]. Figure 1 shows hyperlink network consisting of all these 11 alternative shortest paths, each hyperlink supplied with numbering. Each of 11 alternative shortest paths can be seen when chaining these following 11 series of three hyperlinks: 1, 8, 19; 2, 9, 19; 3, 10, 21; 4, 11, 22; 4, 12, 21; 4, 13, 24; 5, 14, 19; 6, 15, 21; 6, 16, 24; 7, 17, 21; and 7, 18, 19. To depict semantic relationship of each pair of hyperlinked concepts, for each hyperlink we extracted a *relation statement* from sentence surrounding the departing hyperlink in article text, or if suitable sentence was not available then based on other contextual text segments relatively near the hyperlink anchor or possibly based on relation statement we created for another hyperlink going into opposite direction. For example relation statement for hyperlink going from concept Tourism in Malta to concept Economy of Malta is "forms about 15 percent of" which should be interpreted so that "Tourism in Malta forms about 15 percent of Economy of Malta".

We carried out an exploration experiment with a group of 34 students by comparing two learning cases by asking an experiment group (n=24) and a control group (n=10) to perform an exploration task. Students had ages ranging from 15 to 19 years and learning abilities that can be considered normal. We tried our best to form experiment group and control group so that they share approximately same background characteristics, especially in respect to the student's age and self-reported answers how easy it is for him to adopt knowledge through reading and how successfully he performs at school.
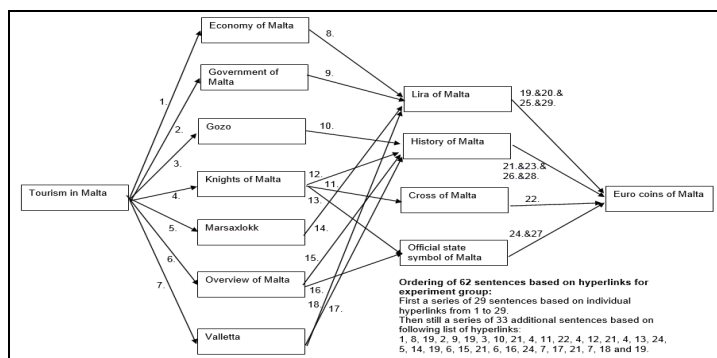
Figure 1. 11 alternative shortest paths having length of three hyperlinks leading from Tourism in Malta to Euro coins of Malta and the ordering of the series of 62 sentences of hyperlinks for experiment group.

Experiment

During one day both members of experiment group and control group participated in three separate learning sessions and each of these sessions consisted of a pre-test, a silent reading task and a post-test based on same fundamental knowledge structure although presented to students in somewhat varied form. For both experiment group and control group the silent reading task in each of three learning sessions was to read once through a series of 62 sentences built by using 22 unique hyperlinks that form the eleven shortest paths leading from concept Tourism in Malta to concept Euro coins of Malta in hyperlink network of Wikipedia category Malta. Each of 22 unique hyperlinks were presented as a sentence consisting of its start concept, relation statement and end concept.

For members of the *experiment group* (n=24) the series of 62 sentences was made to be identical for each of three learning sessions. Here 62 sentences were chained in such an ordering that corresponds to traversing cumulatively a series of associative trails leading from concept Tourism in Malta to concept Euro coins of Malta along alternative parallel shortest paths in hyperlink network of Wikipedia category Malta. Figure 1 illustrates the ordering of the series of 62 sentences for experiment group. So the idea of this ordering is to first introduce the first hyperlink step for each of 11 shortest paths leading from concept Tourism in Malta to concept Euro coins of Malta, next to introduce the second hyperlink step for each of 11 shortest paths, then to introduce the third hyperlink step for each of 11 shortest paths, and after that

[1] To make concepts more easy to understand for students we transformed spelling of some concepts: Maltese euro coins became Euro coins of Malta, Coat of arms of Malta became Official state symbol of Malta, Maltese cross became Cross of Malta, Maltese lira became Lira of Malta, Knights Hospitaller became Knights of Malta and Outline of Malta became Overview of Malta.

finally to introduce one by one the full routes of each of 11 shortest paths (thus showing three consecutive hyperlink steps belonging to each of 11 shortest paths). In contrast with the experiment group, for members of the *control group* (n=10) the series of 62 sentences was made to have randomized ordering of sentences for each of three learning sessions. The idea of this randomization is to enable comparison of control group with experiment group which (as just explained) becomes in each three learning sessions exposed to the series of 62 sentences in such chained ordering that corresponds to traversing cumulatively a series of associative trails leading from concept Tourism in Malta to concept Euro coins of Malta along alternative parallel shortest paths in hyperlink network.

In each of three learning sessions before and after the silent reading task each member of both experiment group and control group was asked to fill in a multiple-choice questionnaire that measured recall of relation statements for each of 22 unique hyperlinks that form the eleven shortest paths leading from concept Tourism in Malta to concept Euro coins of Malta in hyperlink network of Wikipedia category Malta. These pre-tests 1-3 and post-tests 1-3 (that we also call as measurements 1-6) contained always the same 22 multiple-choice items, each item corresponding to each of 22 unique hyperlinks and having four alternative answers we had created so that only one of them is correct. For example a multiple-choice item corresponding to hyperlink Tourism in Malta → Economy of Malta had four alternative answers: Tourism in Malta forms about 10 percent of Economy of Malta; Tourism in Malta forms about 12 percent of Economy of Malta; Tourism in Malta forms about 15 percent of Economy of Malta; Tourism in Malta forms about 17 percent of Economy of Malta (here the third alternative is correct). We created three wrong alternatives for each multiple-choice item by using relatively basic vocabulary with an aim to make them sufficiently confusing with the correct alternative.

Each of six multiple-choice questionnaires given during exploration task have different randomized ordering for 22 multiple-choice items (but in each multiple-choice item the four answer alternatives and their ordering always remain the same in each of six questionnaires). For both experiment group and control group the randomization of ordering of multiple-choice items in six questionnaires of pre-tests and post-tests aims to prevent emergence of such repeated chained ordering of multiple-choice items that might start to compete with the enforcing repetitions occurring in silent reading task of experiment group which corresponds to cumulatively traversing a series of associative trails in hyperlink network. We carried out exploration task concurrently with experiment group and control group so that the first learning session was in the morning lasting from 10:00 to 10:20 (i.e. lasting 20 minutes), the second learning session was in the afternoon lasting from 16:30 to 16:50 (i.e. 6 hours 30 minutes after the first session and lasting 20 minutes) and the third learning session was in the evening lasting from 18:40 to 19:00 (i.e. 8 hours 40 minutes after the first session and lasting 20 minutes).

Based on the ratio of correct answers to wrong answers in multiple-choice questionnaires, Table 1a shows for both pre-test and post-test of each of three learning sessions the degree of recall of relation statements for each of 22 unique hyperlinks that form the eleven shortest paths and also for each of unique 22 pairs of consecutive hyperlinks that exist along the eleven shortest paths. Here it appears that for both unique hyperlinks and consecutive pairs of hyperlinks in the first measurement control group can reach higher degrees of recall than experiment group but after that in the following four measurements experiment group can reach higher degrees of recall than control group (in the sixth measurement for unique hyperlinks experiment group outperforms control group and for consecutive pairs of hyperlinks they share same value of degree of recall).

a)		Pre-test 1	Post-test 1	Pre-test 2	Post-test 2	Pre-test 3	Post-test 3
unique hyperlinks	experiment group	0.35	0.70	0.70	0.80	0.77	0.82
	control group	0.37	0.66	0.61	0.71	0.75	0.79
pairs of consecutive hyperlinks	experiment group	0.14	0.52	0.54	0.68	0.63	0.71
	control group	0.17	0.48	0.43	0.54	0.59	0.71

b)		Pre-test 1	Post-test 1	Pre-test 2	Post-test 2	Pre-test 3	Post-test 3
experiment group in contrast with control group	unique hyperlinks	-0.18298	0.186432	0.40179	0.411672	0.107309	0.165636
	pairs of consecutive hyperlinks	-0.30976	0.187537	0.418071	0.482189	0.181495	0

c)		Pre-test 1	Post-test 1	Pre-test 2	Post-test 2	Pre-test 3	Post-test 3
unique hyperlinks	experiment group	0	1.47	1.43	1.63	1.60	1.63
	control group	0	1.34	1.06	1.36	1.52	1.42
pairs of consecutive hyperlinks	experiment group	0	1.42	1.41	1.59	1.60	1.62
	control group	0	1.17	1.03	1.23	1.43	1.48

Table 1. a) Degree of recall of unique hyperlinks and pairs of consecutive hyperlinks along the shortest paths. b) Effect size in favor of degree of recall in experiment group in contrast with degree of recall in control group. c) Effect size in favor of degree of recall in measurements 2-6 in contrast with degree of recall in measurement 1.

It is positive to note that for both experiment group and control group the degree of recall seems to remain relatively high for quite long period of time even after just one exposure to knowledge and later exposures can further increase degree of adoption. The difference of degree of recall between experiment group and control group seems to be biggest for the third measurement and then for the fourth measurement but after that the difference gets again lower values. For unique hyperlinks interesting is that for experiment group the degree of recall seems to stay constant between measurements 2-3 and for control group the degree of recall seems to rise between measurements 4-5, and for consecutive pairs of hyperlinks interesting is that for experiment group the degree of recall seems to rise between measurements 2-3 and for control group the degree of recall seems to rise between measurements 4-5.

Discussion and future work

In each of three learning sessions, the exploration task seems to offer a promising timing structure that processes same knowledge structure – although in different forms - in three consecutive phases including about 5 minutes for pre-test, about 10 minutes for silent reading task and about 5 minutes for post-test. We think that this our proposed timing practice has positive resemblance with several previous findings, including Harvey and Svoboda (2007) showing that when a spine of synapse is stimulated to action potential also surrounding spines are more sensitive for stimulus for about 10 minutes, Kandel (2001) showing that stimulation of synapses can be successfully triggered by 4–5 spaced puffs of serotonin leading to activation of genes establishing long-term memory, Fields (2005) showing that to activate a gene for long-term memory formation in a synapse there is a need for at least three action potentials at least 10 minutes apart and once the gene is activated it produces required proteins for about 30 minutes, and Tambini (2010) showing that during a rest following an associative encoding task the hippocampal-cortical correlations can predict later associative memory.

Since the effect that an intervention has on learning achievements of students has been often measured with effect size, we wanted to measure the effect size of exploration task in favor of experiment group in contrast with control group and also the effect size along learning of experiment group and along learning of control group. Based on 815 meta-analyses the average of effect sizes - which was 0.40 - has been suggested to be used as a benchmark between effects that need more consideration and effects that are worth having, and it has been suggested that teachers can accomplish on average an effect size of 0.20–0.40 on the student's school achievement per year (Hattie 2009). According to Hattie (Hattie 2009) effects sizes in range 0–0.15 correspond to developmental effects that can be achieved even without schooling, effect sizes in range 0.15–0.40 correspond to effects from a teacher in a typical year of schooling and effect sizes above 0.40 correspond to effects of influences that can be expected to have the greatest impact on the student achievement outcomes. It has been shown that the average effect sizes of six main categories of influences (contributors) to learning according to Hattie's synthesis of 815 meta-analyses are in range 0.23–0.49 (Hattie 2009).

We estimated the effect size in favor of degree of recall in experiment group in contrast with degree of recall in control group by computing the difference of averages of degree of recall in experimental group and control group divided by square root of pooled variance (see Table 1b) for both unique hyperlinks and consecutive pairs of hyperlinks. We got estimates that based on previous research of Hattie (2009) the effect sizes in measurements 3-4 are over 0.40 corresponding to developmental effects with the greatest impact (worth having), and the effect sizes for hyperlinks in measurements 2 and 6 and for consecutive pairs of hyperlinks in measurements 2 and 5 belong to range of 0.15–0.40 corresponding to developmental effects from a teacher in a typical year of schooling. Furthermore we estimated the effect size in favor of degree of recall in measurements 2-6 in contrast with degree of recall in measurement 1 by computing the difference of averages of degree of recall in measurements 2-6 and measurement 1 divided by square root of pooled variance (see Table 1c) for both unique hyperlinks and consecutive pairs of hyperlinks. We got estimates that based on previous research of Hattie (2009) in measurements 2-6 the effect sizes are over 0.40 corresponding to developmental effects with the greatest impact (worth having).

Motivated by our results we think that future research should develop methods for easy generation of educational content based on network models and for guided exploration in network so that knowledge entities can be adopted in a pedagogically rewarding chained sequence.

References

Akbulut, Y., & Cardak, C. (2012). Adaptive educational hypermedia accommodating learning styles: a content analysis of publications from 2000 to 2011. Computers & Education, 58(2), 835-842.

Lahti, L. (2014). Educational exploration based on conceptual networks generated by students and Wikipedia linkage. Proc. World Conference on Educational Multimedia, Hypermedia and Telecommunications (EdMedia 2014), 964-974, AACE. ISBN 978-1-939797-08-7. http://www.editlib.org/p/147608/

Lahti, L. (to appear). Computer-assisted learning based on cumulative vocabularies, conceptual networks and Wikipedia linkage. Doctoral dissertation (submitted for evaluation in January 2014), Department of Computer Science and Engineering, Aalto University School of Science, Finland. vttbza vcy pwjxjy ul ölvvlcopy alty vzvpl, vltvvpl sgdjj bztdzy bptwwp dlöåc öltxz bccwt åjtdt lyul öttvvl ul vltvvt vzsbllxlyt.

Bellissens, C., Jeuniaux, P., Duran, N., & McNamara, D. (2010). A text relatedness and dependency computational model: using latent semantic analysis and Coh-Metrix to predict self-explanation quality. Studia Informatica Universalis, 8, 85-125.

Bouchet, F., Harley, J., Trevors, G., & Azevedo, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. Journal of Educational Data Mining, 5(1).

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction, 6(2-3), 87-129.

Chen, C., & Chen, Y. (2009). Effectiveness of constructed responses and multiple-choice questions on recall and recognition in a web-based language learning environment. Proc. 17th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education, Hong Kong, China.

Chen, S. (2002). A cognitive model for non-linear learning in hypermedia programmes. British Journal of Educational Technology, 33 (4), 449-460.

Fields, R. (2005). Making memories stick. Scientific American, 292 (February 2005), 74-81.

Gureckis, T., & Markant, D. (2012). Self-directed learning: a cognitive and computational perspective. Perspectives on Psychological Science, 7(5), 464–481.

Hattie, J. (2009). Visible learning: a synthesis of over 800 meta-analyses relating to achievement. Routledge, London, UK.

Harvey, C., & Svoboda, K. (2007). Locally dynamic synaptic learning rules in pyramidal neuron dendrites. Nature, 450(7173).

Kandel, E. (2001). The molecular biology of memory storage: a dialog between genes and synapses. Nobel Lecture, 8 December 2000. Bioscience Reports, 21(5).

Kastner, M., & Stangl, B. (2011). Multiple choice and constructed response tests: do test format and scoring matter? Procedia - Social and Behavioral Sciences, 12, 263-273.

Liao, C., Chen, Z., Cheng, H., & Chan, T. (2012). Unfolding learning behaviors: a sequential analysis approach in a game-based learning environment. Research and Practice in Technology Enhanced Learning, 7(1), 25-44.

Martin, B., Mitrovic, T., Mathan, S., & Koedinger, K. (2011). Evaluating and improving adaptive educational systems with learning curves. User Modeling and User-Adaped Interaction, 21, 249–283.

Masucci, A., Kalampokis, A., Eguíluz, V., & Hernández-García, E. (2011). Wikipedia information flow analysis reveals the scale-free architecture of the semantic space. Public Library of Science ONE, 6(2), e17333.

Mohamed, H., Bensebaa, T., & Trigano, P. (2012). Developing adaptive intelligent tutoring system based on item response theory and metrics. International Journal of Advanced Science and Technology, 43.

Robberecht, R. (2007). Interactive nonlinear learning environments. The Electronic Journal of E-Learning, 5(1), 59-68.

Soh , L., & Blank, T. (2008). Integrating case-based reasoning and meta-learning for a self-improving intelligent tutoring system. International Journal of Artificial Intelligence in Education, 18 (1), 27-58.

Tambini, A., Ketz, N., & Davachi, L. (2010). Enhanced brain correlations during rest are related to memory for recent experiences. Neuron, 65(2), 280-290.

Thalmann, S. (2014). Adaptation criteria for the personalised delivery of learning materials: a multi-stage empirical investigation. Australasian Journal of Educational Technology, 30(1).

Thompson, G., Kello, C., & Montez, P. (2013). Searching semantic memory as a scale-free network: evidence from category recall and a Wikipedia model of semantics. Proc. 35th Annual Meeting of the Cognitive Science Society. Cognitive Science Society, Austin, TX, USA, 3533-3538. ISBN 978-0-9768318-9-1.

Thomson, K., Watt, A., & Liukkonen, J. (2014). Developmental and cultural aspects of field-dependence in 11 and 12 year old Estonian and Finnish students. TRAMES - A Journal of the Humanities and Social Sciences, 18(68/63), 1, 89-101.

Tintarev, N. (2009). Explaining recommendations. Doctoral dissertation, Department of Computing Science, University of Aberdeen, Scotland, UK.

Yen, J. (1971). Finding the k shortest loopless paths in a network. Management Science, 17(11), 712-716.