A Semantic Recommender System for Adaptive Learning

**Paolo Montuschi, Fabrizio Lamberti, Valentina Gatteschi,
and Claudio Demartini,** *Politecnico di Torino*

Individuals must continuously update their qualification levels to stay relevant in today's market. The authors' semantic-based recommender system crosses heterogeneous information about individuals' backgrounds and advertised jobs with an online course catalog to identify appropriate learning resources.

Today's economic and technological landscape emphasizes the importance of acquiring the right competencies to successfully enter the labor market. Having obtained these competencies, individuals must continuously update them to stay relevant on the job. For example, the time it takes for half the knowledge of an IT professional to be superseded is now just 10 to 12 years.[1]

Given this scenario, all stakeholders face crucial problems every day, including questions of how

- learners and job-seekers can plan their learning to ensure that the competencies they acquire will largely match those the labor market requests;

- companies can schedule on-the-job training activities to keep their work forces up-to-date; and

- education and training providers can tailor their offers to both companies' and learners' evolving needs.

Traditionally, such concerns were dealt with by a mentor—that is, someone with considerable experience who could see the whole picture, had perspective (about future developments), and possessed in-depth knowledge about the constraints, weaknesses, or strong points of an applicant. This mentor's job was to compare information about learners' or job-seekers' backgrounds and the company's requirements with education and

training offers to find the best match. However, in a learning scenario that's getting ever more globalized and dematerialized,[2,3] the increasing amount of heterogeneous data that must be considered is making it more difficult and time-consuming to process such information.

Fortunately, several technologies already exist, from basic keyword filters to highly sophisticated tools, that let machines intervene in this process. Based on these technologies, automatic recommender systems can be created that can complement the essential role human decision makers play by reducing data's size and complexity and letting humans focus on the real and most relevant issues.

In this context, we demonstrate how a recommender system for adaptive learning and mentoring can be implemented that uses emerging semantic computing technologies. Our prototype system exploits semantics to analyze various lifelong learning-related data—such as job-seekers' résumés, worker profiles, job offers, and company (re)training requirements—and match this data with a catalog of available courses to find (and suggest) those that could address specific deficits.

To our knowledge, this is the first work that integrates the concerns of job-seekers and workers, companies, and education providers. By eliciting possible competency gaps and identifying suitable remedies, our system can contribute to the shift from a one-size-fits-all paradigm to a new approach to learning in which required or relevant courses are retrieved based on (that is, adapted to) learners' needs and characteristics, thus providing concrete answers to the critical questions we've posed.

Recommender Systems

Since their first appearance in the mid-1990s, recommender systems have gained attention from both research and commercial fields.[4] Today, they are widely used in several applications to recommend goods, as in the case of Amazon (www.amazon.com); movies and songs, as with MovieLens (www.movielens.umn.edu), Pandora (www.pandora.com), and Last.fm (www.last.fm); and connections, as with LinkedIn (www.linkedin.com) or Facebook (www.facebook.com).

These systems aim to compute a rating for any given item in the domain of interest and provide users with those items that obtain the highest rating according to users' characteristics. We can classify recommender systems based on the following three categories:

- *Content-based*. Recommendations are made on the basis of explicit information (users' evaluation of items, forms they've filled out, and so on) or implicit information (users' past behavior). Being based on past ratings or actions, such systems risk recommending items that are too similar to those the user previously considered. Moreover, content must be expressed in a format that enables automatic processing.
- *Collaborative*. Recommendations are based on the behaviors and ratings of similar people. These systems compute similarity among users and make rating predictions by combining evaluations of a person's nearest neighbors

> Even though recommender systems are successfully exploited in several contexts, their application to the learning scenario requires more than a simple adaptation.

(memory-based algorithms) or creating a model based on available ratings (model-based algorithms). A key limitation is that new items or those that have been rated by only a few people are rarely suggested.
- *Hybrid*. Recommendations are made by combining content-based and collaborative approaches.

Even though recommender systems are successfully exploited in several contexts, their application to the learning scenario requires more than a simple adaptation.[5] In fact, the recommendation of learning resources should be done based on a wider set of variables, including learners' knowledge level, learning goals, and cognitive style.[6] With this view, several approaches have been proposed and applied to the learning context:

- *Bayesian networks*—probabilistic models representing random variables (nodes) and conditional dependencies (edges) in a directed acyclic graph. In the learning scenario, they have been applied to help infer which topics are better known by a given student to provide that student with suggestions on aspects for further revision or study.[7]
- *Association rules*—a research method for discovering correlation among items in large datasets. This technique, which is usually exploited in business trade contexts (for instance, to compute the probability that if a customer buys item *X*, he or she will also buy item *Y*), has been used to guide learners while they browse online resources by easing the identification of frequent patterns in the best learning strategies.[8]
- *Clustering*—a technique for grouping elements into classes by maximizing the similarity

> **Our goal was to investigate whether semantics could be exploited in a more integrated scenario, encompassing lifelong learning and training personalization.**

among those belonging to the same group. Clustering has been applied to learning objects to improve retrieval[9] or to learners to group them on the basis of their behavior.[10]
- *Genetic algorithms*—an optimization technique for searching a function's global optimum by starting from randomly generated solutions and iteratively combining or varying them. Such algorithms have been used for appropriate personalized curriculum sequencing based on learners' mastery level.[11]
- *Semantics*—an approach for defining machine-understandable resources. Using semantics, machines could go beyond pure keyword-based processing to understand natural language and reason in a way similar to human thinking. Semantics have been used to identify similarities among concept maps created by learners,[12] recommend learning resources on

the basis of their tags and number of hits,[13] and analyze training material characteristics and match them with users' interests.[14]

Because our objective was to compare information expressed in natural language coming from multiple sources and domains, we identified semantics as the most suitable approach for our system. This choice was substantiated by extensive research results on semantics applied to education (see www.computer.org/portal/web/computingnow/archive/april2014).

Semantics-Based Course Recommender System

We'd already successfully used semantics-based technology in the MATCH project (http://match.cpv.org), which aimed to automatically compare and match résumés and job postings to support job seeking and recruiting processes. Our new goal was to investigate whether semantics could be effectively exploited in a broader and more integrated scenario, encompassing lifelong learning and training personalization.

We embedded our recommender system into the LO-MATCH platform (www.lo-match.polito.it) developed for the MATCH project; this closed the loop and provided users with a powerful tool for identifying their competency gaps and suggesting suitable actions for filling these gaps.

Before analyzing the technical choices we made, let's look at the main steps necessary for the development of any semantics-based system.

First, the domain of interest should be described by creating formal models to represent it (*taxonomies*, or hierarchical structures of terms, or *ontologies*, "formal, explicit specifications of a shared conceptualization"[15]). This is often a time- and effort-consuming manual step because it requires that domain experts identify relevant concepts and relate them to each other. An alternative to creating new models could be to exploit existing taxonomies and ontologies, or to semi-automatically build them (although result validation could be cumbersome[16]).

Next, information for processing should be described, or "annotated," by linking it to concepts defined in the aforementioned taxonomies or ontologies. Annotation could be carried out manually (with significant effort, even for skilled operators), or (semi-)automatically (with less accurate results).

Finally, suitable algorithms for computing the semantic distance between different concepts (by browsing relations among them) should be created.

As we describe in prior work,[17] in developing the initial LO-MATCH platform, we made the following key design choices.

To create the ontology, we adopted a hybrid strategy by starting from an already existing ontology and letting the user extend it via a graphics tool. Specifically, we chose the WordNet semantic thesaurus,[18] an extremely rich ontology that organizes words in synsets (sets of synonyms) linked through semantic relations.

To perform annotation, we designed an innovative drag-and-drop interface that lets users intuitively pick the most suitable synsets from the ontology to express the content of a course, competency, and so on. The system was also designed to work without explicit annotations, even though, in such cases, comparisons would be based on all the possible meanings of terms contained in the resources to be compared.

Instead of relying on the general concept of competency, as is done in similar works, we exploited the guidelines defined by the European Qualification Framework (http://ec.europa.eu/eqf) and focused on the concept of learning outcome, intended as "a statement of what a learner knows, understands, and is able to do on completion of a learning process" (further detailed as *knowledge*, *skill*, and *competence* elements). Thus, résumés, job offers, and learning resources were expressed via a well-known standard shared by actors from both education and labor.

Starting from this framework, we developed our recommender system. The process starts with the identification of the competency gap between possessed and required learning outcomes, which is found by matching a learner's profile or job seeker's résumé with a job offer posted by a given company. Then, the system queries a semantic-aware catalog of learning resources to recommend the courses that could possibly fill that gap. In particular, our system can

- carry out a semantic comparison based on words used for expressing the competency gap and the course content (a number of parameters allow the user to specify whether to give more relevance to concepts found in the title, the description, and so on);

- support the annotation of courses via a tagging-like operation based on the same catalog of learning outcomes used to describe résumés and job postings (to make the comparison even more precise); and

- consider language requirements (expressed as listening, reading, spoken interaction, spoken production, and writing abilities) while determining which courses to recommend.

To test our system, we linked it to a local installation of the Moodle learning management system (https://www.moodle.org). This way, trainers and learners would have a way to store and access learning objects, give and receive grades, and so on. Nevertheless, the designed system is independent of the final instruction delivery.

Compared to other learning-oriented recommender systems, our prototype has several advantages:

> **Annotation is easier to perform because users can receive a list of learning outcomes that are semantically related to terms they're using for tagging.**

- It can manage both labeled (that is, annotated or tagged) and unlabeled (expressed in natural language) content.
- Annotation, if required, is easier to perform because users can receive a list of learning outcomes that are semantically related to terms they're using for tagging.
- It hides the complexity derived from the use of semantics by providing a simple interface suitable for nonskilled users.
- It could be theoretically used in different contexts because it relies on a general ontology.
- It relieves users from the burden of specifying their competency gap, which is automatically computed.
- It doesn't build onto the knowledge of other learners' behavior.
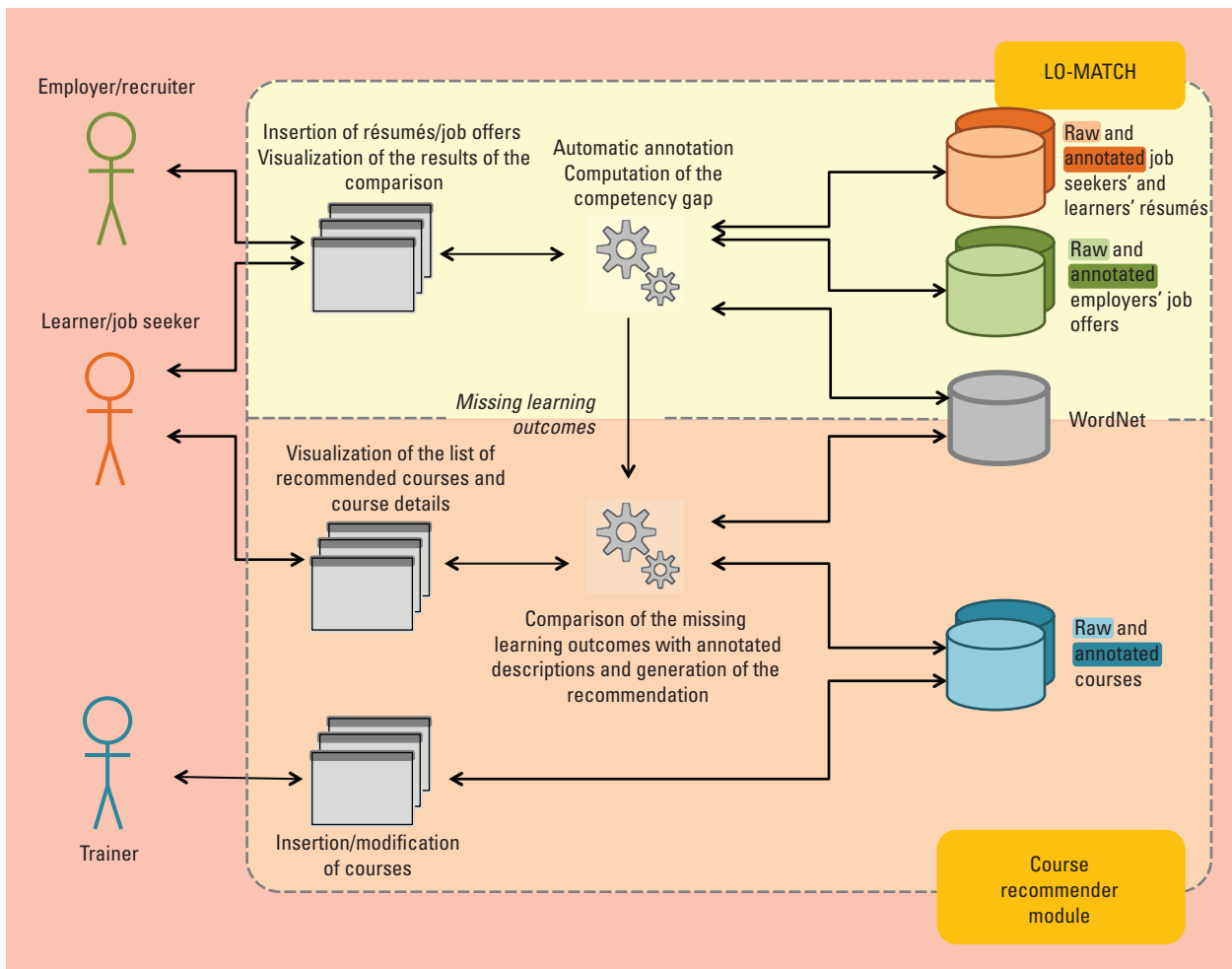- It incorporates European standards.

Figure 1. Architecture and main features of the proposed recommender system. The LO-MATCH system computes the competency gap in terms of missing knowledge, skill, and competence elements.[17] The recommender system module cross-references missing learning outcomes with the semantically annotated courses and suggests those that could let users fill their gaps. (In our implementation, the catalog of courses is stored in a Moodle instance.)

- It's the first system (to the best of our knowledge) that integrates job seeking, employment, and training contexts.

Figure 1 shows the system's architecture. A Web-based front end can be used to insert résumés, job offers, and courses, which are stored in their native (raw) format. On the back end, the semantic logic is responsible for automatically annotating inserted resources, supporting the user in possible refinement steps, and recording semantics-aware information in dedicated knowledge bases (the use of distinct repositories would ease the process of linking the system to other résumés, job offers, or course catalogs).

The LO-MATCH matchmaking engine computes the competency gap in terms of missing knowledge, skill, and competence elements.[17] Finally, the recommender system module cross-references missing learning outcomes with the semantically annotated courses and suggests those that could allow users to fill competency gaps.

A video demonstration of our system is available at http://youtu.be/A8c5MwnlGvA.

A Practical Example

To better understand how the proposed recommender system functions, consider the example in Figure 2.

Alice is a job seeker from the UK who is *able to interact with customers* and *prepare cocktails*. Alice
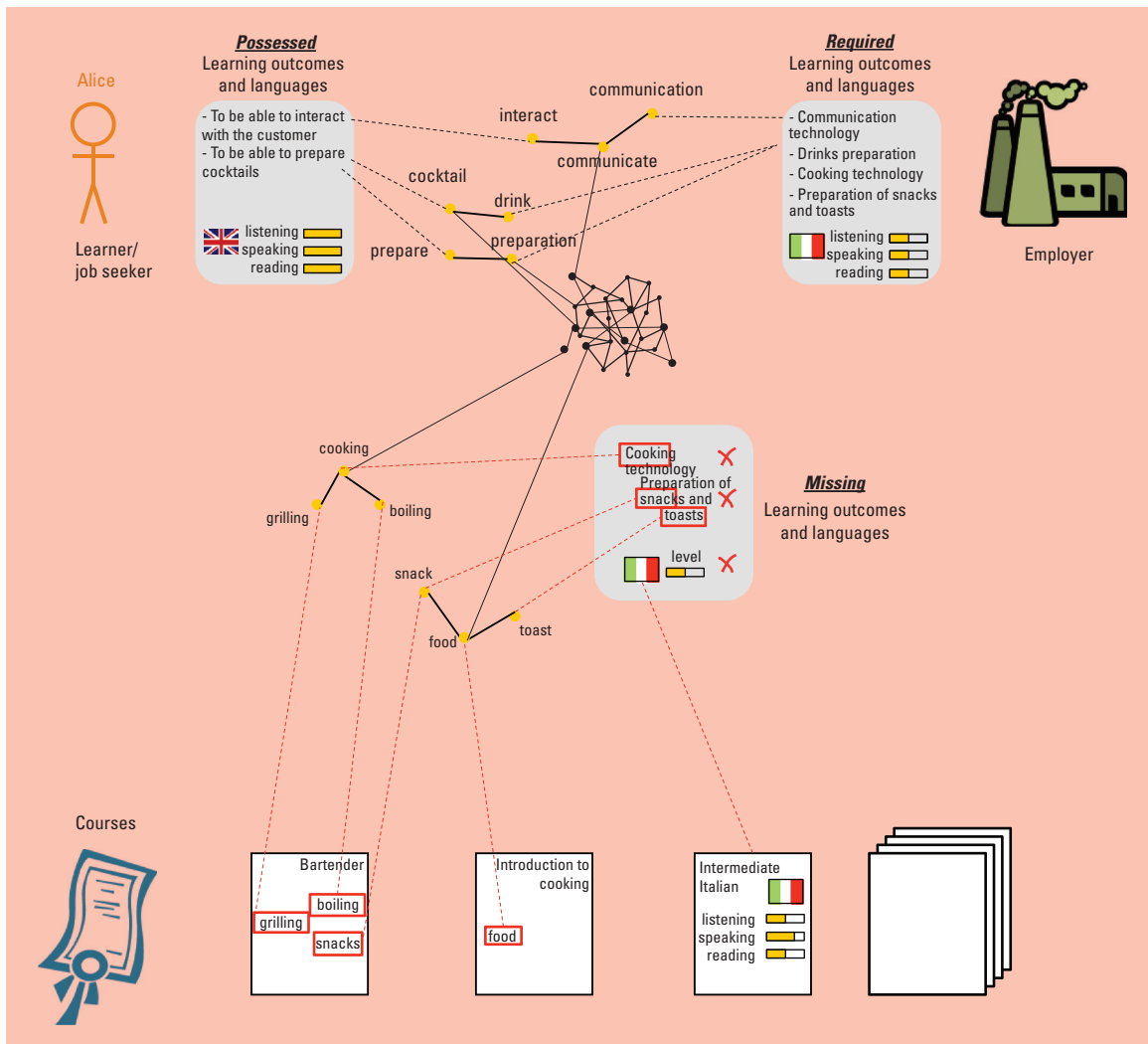
Figure 2. Recommending the best courses based on the gap between possessed and requested learning outcomes. Course rankings are computed by semantically comparing the sentences used in the learner's or job-seeker's résumé with the company's requirements and then with course descriptions.

compares her résumé with the company's requirements using LO-MATCH (upper part of the figure). She finds out that she lacks the *cooking techniques* and *preparation of snacks and toasts* learning outcomes that the company requires, but possesses required learning outcomes somehow related to *communication techniques* (because of the relation *interact-communicate-communication*) and *drinks preparation*; moreover, she lacks the intermediate knowledge of *Italian language* (central part of the figure).

The recommender system compares missing learning outcomes with the course catalog and discovers a *bartender* course that contains concepts related to *boiling, grilling, and snacks*, as well as an *introduction to cooking* course that teaches general subjects about *food* (lower part of the figure). Based on these findings, the system would suggest the *bartender* course (because it provides learning outcomes about *grilling and boiling*, particular types of *cooking techniques*, and *snacks* knowledge), whereas the other course would receive a lower rank. Regarding languages, recommended courses are identified by comparing the average requested proficiency levels with the provided ones.

The comparison between missing learning outcomes and course content is performed as follows:
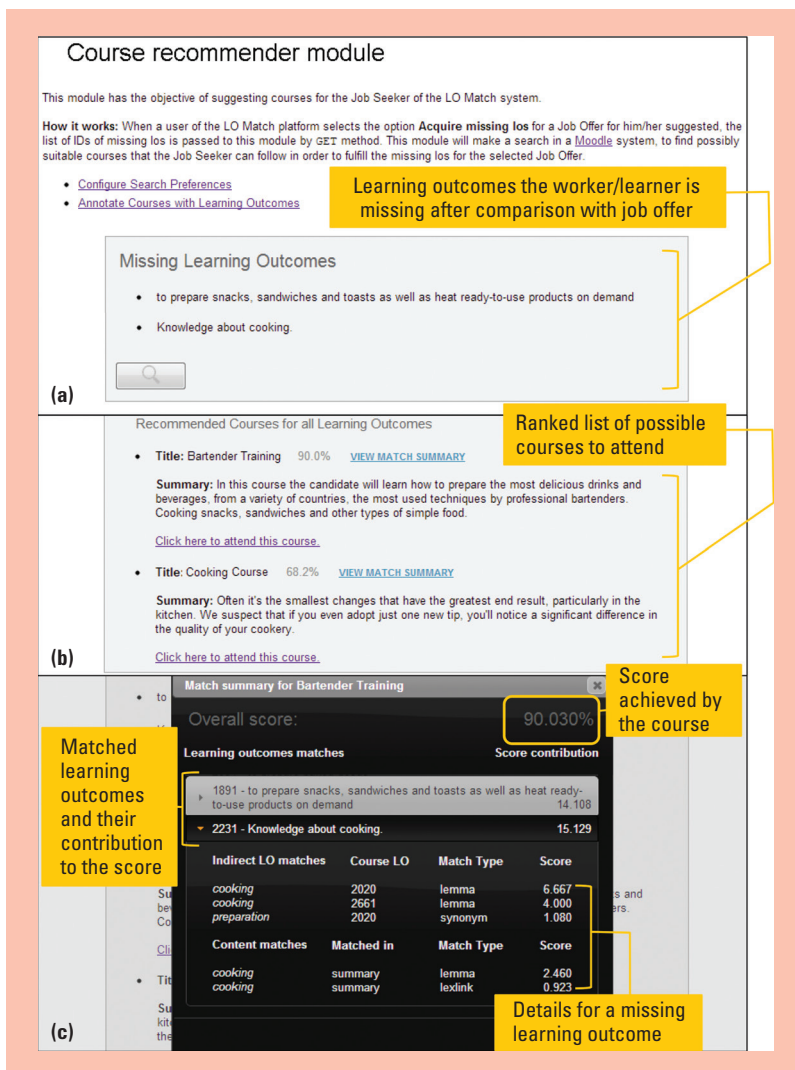
Figure 3. Recommender system interface. Users can see their (a) missing learning outcomes as returned by LO-MATCH, (b) a ranked list of recommended courses, and (c) details for the rank of a given course.

with searched terms (for example, synonyms). To avoid having the comparison biased by the richness of the descriptions, the overall number of words is also considered. Should the course have been previously tagged via a set of learning outcomes, a similar approach based on the number of matched terms and the type of semantic relations would be adopted. The contribution of these parameters has been fine-tuned using validation results collected in our former project initiative.

Graphics Interface

Figure 3 shows some screenshots of the designed system. When learners or job seekers receive their competency gap, they can use the course recommender module shown in Figure 3a. Here, users can search for courses that provide the learning outcomes they're missing. The system presents the matching courses as a ranked list (Figure 3b). Each course is accompanied by a score (percentage) based on how much it satisfies users' needs. Should users want to further investigate the reason behind a given result, they can click on "view match summary." A dialog window (3c) reports details about how each missing learning outcome was matched and contributed to the final rank.

Experimental Tests

Existing recommender systems were exploited in domains that were different from the one tackled here or, when used in the education field, that didn't provide the functionalities of our prototype platform. Therefore, an established benchmark isn't available, and a direct comparison with alternative approaches isn't presently feasible.

Hence, for testing purposes, we populated our course catalog with 15 courses from the food and drink, automotive, and IT sectors. Syllabi were taken from national qualification repositories, sectoral standards, or other free online databases. A panel of 10 trainers verified the effectiveness of the recommendation computed by the semantic system.

- Sentences used in missing learning outcomes and course descriptions are preprocessed to delete stop words and lemmatize terms.
- Each word belonging to learning outcomes is then compared with terms contained in the courses.
- Words that have a given semantic relation (synonyms, hypernyms, hyponyms, and so on) with lemmatized terms are also matched.

The result of the comparison would then be a function of the following parameters: the number of matched words, the portion of text in which they've been found (title, summary, and so on), and the type of relationship they have

We ran experiments using three hypothetical job-seekers, each looking for a job in one of the aforementioned sectors. Trainers were provided with a booklet containing course descriptions and missing learning outcomes for the three job seekers, and were asked whether they would recommend a given course by choosing among *definitely no* (1), *more no than yes* (2), *more yes than no* (3), and *definitely yes* (4). Scores computed by the system were mapped onto a 1–4 scale as well, based on their quartile.

The booklet and the results of the automatic recommendation are available as a Web extra at http://doi.ieeecomputersociety.org/10.1109/MITP.2015.75.

The agreement between manual and automatic recommendations (linearly weighted Cohen's $\kappa = 0.796$) confirmed the validity of the approach. For instance, in the IT scenario, the system succeeded in suggesting courses dealing with website development and server-side scripting (as required by missing learning outcomes), and discarded courses on mobile applications or networking.

Technological and social changes are making jobs in the medium- and long-term future hard to predict with reliability. In this context, finding the right courses to attend and being able to adapt to the labor market's needs on-the-fly are of paramount importance.

Experimental results for our system confirmed that semantic technology could represent a valid solution to prefilter the large amount of data in job seekers' résumés, workers' profiles, job offers, and companies' (re)training requirements and compare the results with available learning resources.

Even though the interface to our proposed recommender system has been designed for learners, its use could easily be extended to cope with the needs of other users while keeping the same processing logic. For instance, it could support training providers in the identification of innovation needs coming from the labor sector, which could be used to drive the design of new courses. It could be exploited by companies as a tool for planning on-the-job training. Professional societies could offer it as a service to improve their customer loyalty strategies. Note that the advantages of semantic processing will continue to grow as the amount of data available for analysis increases, thus making the benefits in terms of time and money savings ever more evident.

Future work could be devoted to adding these functionalities by integrating information coming from existing sources (job portals, training offers published online, companies' competency databases, and so on). Moreover, at present, the system can't identify proficiency levels for learning outcomes. Hence, future research activities could investigate to what extent the different components of a learning outcome convey information about its proficiency, with the aim of developing automatic techniques that can recommend the best courses based on the relationships between more specific and more general topics.

We strongly hope that, in the future, the domain of adaptive education will receive increased attention from researchers as well as significant investments to enable the development of instruments that, apart from having a considerable social impact, could be used to train learners for future professional needs. 🖽

References

1. S. Murugesan, "Succeeding as an IT Professional," *IT Professional*, vol. 16, no. 1, 2014, pp. 2–4.
2. D.R. Garrison, *E-Learning in the 21st Century: A Framework for Research and Practice*, Taylor & Francis, 2011.
3. M. Gaebel, *MOOCs—Massive Open Online Courses*, EUA Occasional Papers, 2013.
4. G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Trans. Knowledge and Data Eng.*, vol. 17, no. 6, 2005, pp. 734–749.
5. H. Drachsler, H. Hummel, and R. Koper, "Personal Recommender Systems for Learners in Lifelong Learning Networks: The Requirements, Techniques, and Model," *Int'l J. Learning Technology*, vol. 3, no. 4, 2008, pp. 404–423.
6. K. Verbert et al., "Context-Aware Recommender Systems for Learning: A Survey and Future Challenges," *IEEE Trans. Learning Technologies*, vol. 5, no. 4, 2012, pp. 318–335.
7. F. Colace and M. De Santo, "Ontology for E-Learning: A Bayesian Approach," *IEEE Trans. Education*, vol. 53, no. 2, 2010, pp. 223–233.
8. Y. Zou, "Personalized Automatic Recommendations for the Web-Based Autonomous Language Learning System Based on Data Mining Technology," *Proc.*

Int'l Symp. IT in Medicine and Education (ITME), 2011, pp. 326–329.

9. A.S. Sabitha and D. Mehrotra, "User Centric Retrieval of Learning Objects in LMS," *Proc. 3rd Int'l Conf. Computer and Communication Technology* (ICCCT), 2012, pp. 14–19.

10. F-H. Wang and H-M. Shao, "Effective Personalized Recommendation Based on Time-Framed Navigation Clustering and Association Mining," *Expert Systems with Applications*, vol. 27, no. 3, 2004, pp. 365–377.

11. M-J. Huang, H-S. Huang, and M-Y. Chen, "Constructing a Personalized E-learning System Based on Genetic Algorithm and Case-Based Reasoning Approach," *Expert Systems with Applications*, vol. 33, no. 3, 2007, pp. 551–564.

12. A.A. Kardan, S. Abbaspour, and F. Hendijanifard, "A Hybrid Recommender System for E-Learning Environments Based on Concept Maps and Collaborative Tagging," *Proc. 4th Int'l Conf. Virtual Learning* (ICVL), 2009, pp. 300–307.

13. S.F. Mohsin and R.U. Rashid, "Web-Based Multimedia Recommendation System for E-Learning Website," *Int'l J. Advanced Networking and Applications*, vol. 1, no. 4, 2010, pp. 217–223.

14. Q. Yang et al., "Semantic Web-Based Personalized Recommendation System of Courses Knowledge Research," *Proc. Int'l Conf. Intelligent Computing and Cognitive Informatics* (ICICCI), 2010, pp. 214–217.

15. T.R. Gruber, "A Translation Approach to Portable Ontology Specifications," *Knowledge Acquisition*, vol. 5, no. 2, 1993, pp. 199–220.

16. I. Bedini and B. Nguyen, *Automatic Ontology Generation: State of the Art*, tech. report, Univ. of Versailles, 2007.

17. P. Montuschi et al., "Job Recruitment and Job Seeking Processes: How Technology Can Help," *IT Professional*, vol. 16, no. 5, 2014; http://doi.ieeecomputersociety.org/10.1109/MITP.2013.62.

18. G.A. Miller, "WordNet: A Lexical Database for English," *Comm. ACM*, vol. 38, no. 11, 1995, pp. 39–41.

Paolo Montuschi *is a professor in the Department of Control and Computer Engineering at Politecnico di Torino. His research interests include computer arithmetic and architectures, computer graphics, electronic publications, semantics and education, and new frameworks for the dissemination of scientific knowledge. Montuschi is an IEEE Fellow and an IEEE Computer Society Golden Core member; serves as editor in chief of* IEEE Transactions on Computers; *and is a member of the* Computing Now *advisory board, the IEEE Publications Services and Products Board, and the IEEE Products and Services Committee. He is a life member of the International Academy of Sciences of Turin. Contact him at paolo.montuschi@polito.it; http://staff.polito.it/paolo.montuschi.*

Fabrizio Lamberti *is an associate professor in the Department of Control and Computer Engineering at Politecnico di Torino. His research interests include computational intelligence, semantic processing, distributed computing, human-computer interaction, computer graphics, and visualization. Lamberti is a senior member of IEEE and the IEEE Computer Society and serves as an associate editor for* IEEE Transactions on Emerging Topics in Computing *and* IEEE Consumer Electronics Magazine. *Contact him at fabrizio.lamberti@polito.it; http://staff.polito.it/fabrizio.lamberti.*

Valentina Gatteschi *is a postdoctoral research assistant in the Department of Control and Computer Engineering at Politecnico di Torino. Her main research interests are in semantics and natural language processing. Gatteschi has been involved in several European projects on education. She received a PhD in computer engineering from Politecnico di Torino. Contact her at valentina.gatteschi@polito.it.*

Claudio Demartini *is a full professor in the Department of Control and Computer Engineering at Politecnico di Torino, where he teaches information systems and innovation and product development. His research interests are in software engineering, architectures, and Web semantics. Demartini is the chair of the Control and Computer Engineering Department and a member of the Academic Senate of Politecnico di Torino as well as a consultant on vocational education and training for the Ministry of University, Research, and Education. He is a senior member of IEEE and the IEEE Computer Society. Contact him at claudio.demartini@polito.it; http://staff.polito.it/claudio.demartini.*