Utilizing educational data in collaboration with industry

Pavel Kordík, Stanislav Kuznětsov

Faculty of Information Technology, Czech Technical University in Prague, Czech Republic
pavel.kordik@fit.cvut.cz, kuznesta@fit.cvut.cz

Abstract. Universities are seldom using their data efficiently. In this case study, we show how educational data can be used to recommend suitable students for project, get feedback from industrial partners, help students to focus on skills that are demanded by companies. We have developed portal for students collaboration with industrial partners and run it in a pilot for almost a year. Based on our observations described in this contribution, we are adjusting the portal to enhance the functionality and streamline processes supported by the portal.

Keywords: Educational data mining, Data transformation, ETL, data warehouse

1 Introduction

Education data, especially evaluation of students are often only stored and not utilized due to confidentiality and security reasons. In this paper we show that these data can bring significant added value in many applications. One of the most important application is matchmaking students and industrial partners. Our portal utilize data warehouse and shows industrial partners aggregated data as well as evaluation of students that agreeing to apply for their assignments.

We designed data warehouse and transform evaluation of students in single courses to skills that are more comprehensible to industrial partners. Such skills can then be used to recommend students that are needed.

2 Related work

There not many applications of data mining in university-industry relation (see [3] or [1]) .

VIVO [2] is an open source semantic web platform originally developed and implemented in the Cornell University, after that in 2009 supported by another five universities in USA; it was extended as a tool able to integrate profiles in many different institutions. This platform allows the discovery of the researchers and the universities technical knowledge in many fields, using the professionals

profiles linked and related information. It uses ontologies and helps to find corresponding researchers using keywords extracted from research articles.

When if comes to students, we do not have many information that can be used to recommend relevant ones. Evaluation from courses are very diverse and we need some ontology to transform data into uniform shape.

ACM [6] developed hierarchical tree of skills in the area of Computing that corresponds well to the focus of our faculty. The part of the tree can be observed in Figure 1.

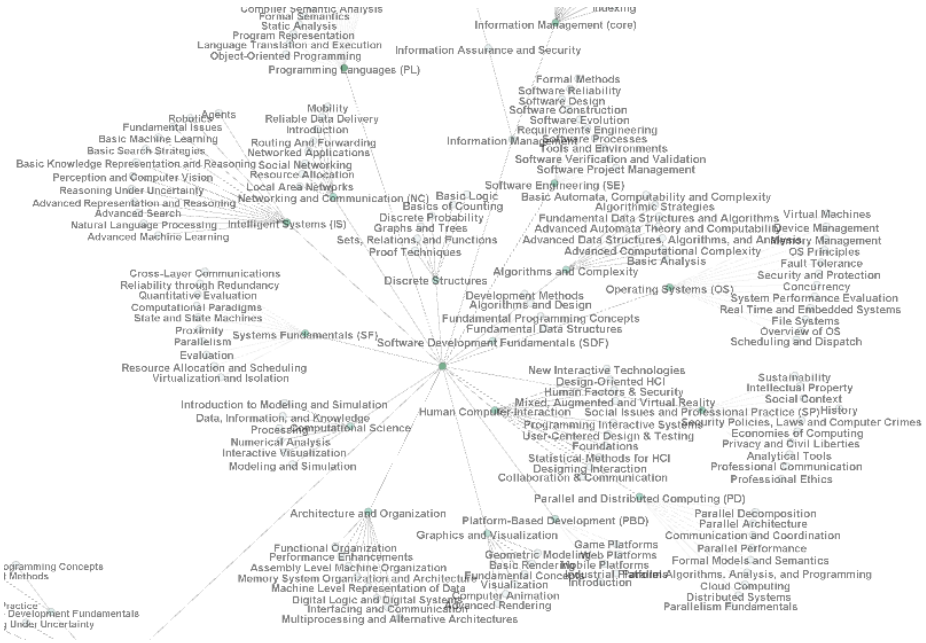

Fig. 1. Map designed by ACM to unify computing curricula.

However such ontology is too complex and not comprehensible to industrial partners of universities.

3 Our approach

Inspired by ACM ontology, we simplified the tree to be more comprehensible for industry partners and better fit to the profile of our faculty. We also did some field research to find out which skills industry partners seeks for and incorporated them to the map (Figure 2).

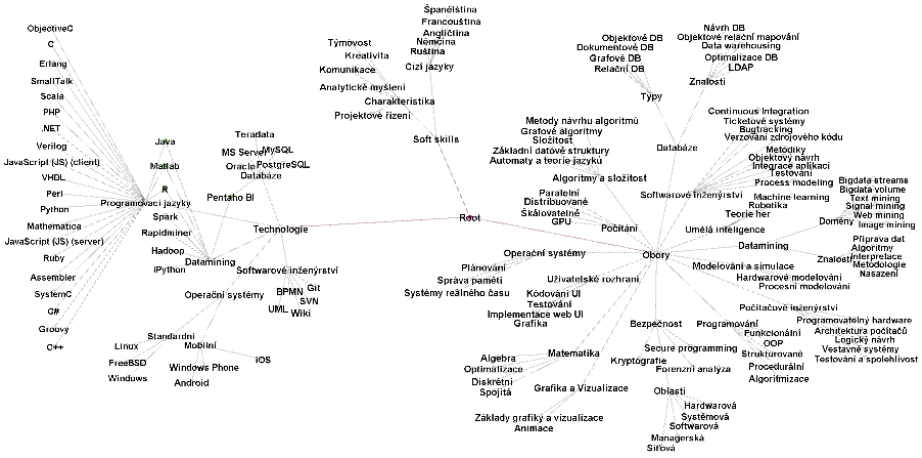

Fig. 2. Redesigned map for the purpose of interaction with industry partners.

3.1 Mapping courses

When the hierarchy of skills was ready, we started to map courses to individual skills. Figure 3 shows how skills are computed from evaluation in individual courses.

Student with better results from a course will get more points to skills that are influenced by the course.

3.2 Getting data

Educational data is very valuable and feed the recommender engine of our system. Utilizing data we build profiles of our students and match them with positions requested by industrial partners. Our faculty uses several information systems to store numerical evaluations of students. E.g. some programming assignments are evaluated using Progtest machine. Such data are hard to interpret by external partners. Therefore we need to consolidate all data in one place in a comprehensible form. For this purpose we designed and build simple data warehouse. At Figure 4, you can see the version 1.0 schema. We inspired by this publication [4] . This version was a pilot and operates since July 2013. To extract data form source systems and feed it into data warehouse, we use several Extract Load Transform (ETL). The transformations are very clear and simple, because all we need is to get the data, mapped it to our data structure and make some calculation to compute students profiles (skills).

Then we implemented another set of ETL operations to compute skills according to ontologies above. These data can be used for various purposes.

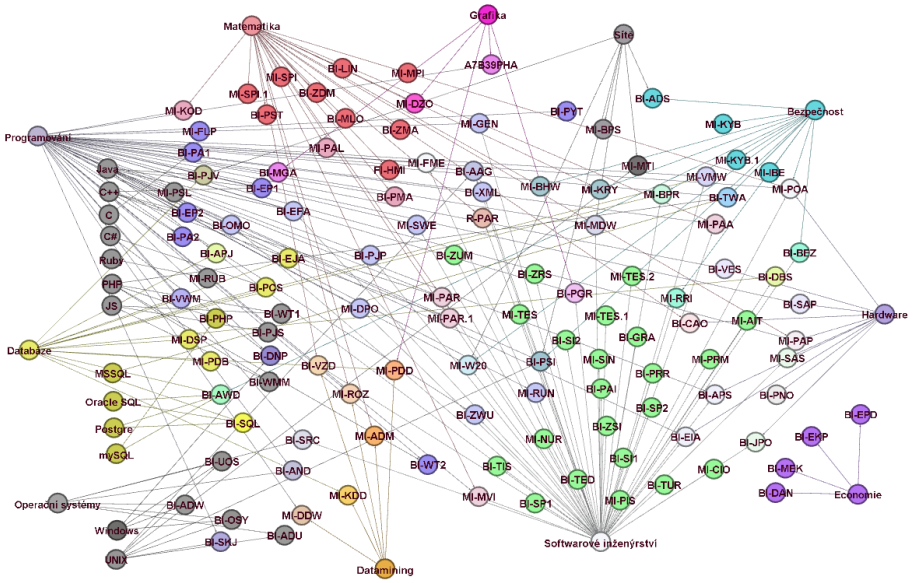

Fig. 3. Evaluation in courses is used to compute skills. Students obtaining better grades receive more points to skills connected with particular course.

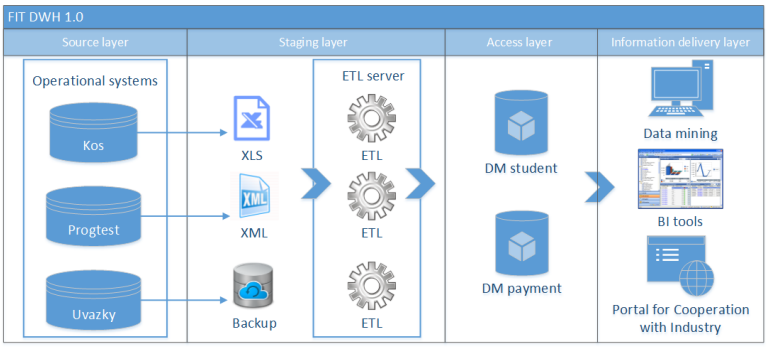

Fig. 4. ETL operations fetch data from educational systems and transforms it to data cubes that can be accessed by target applications.

4 Portal for Cooperation with Industry

Portal is an information system designed, implemented and managed by Faculty of Information Technologies. It provides a communication channel between students and companies to allow them to work together in new projects. This is done by companies exposing their industry needs and specifying the requirements for their goals to be accomplished. With this information the system connects these needs with students who have these specific abilities (skills), allowing a company to select the best candidates to develop the project. By the other side, with these projects students earn experience in real life situations, improve their capabilities by putting their knowledge to the test and they may get paid for the services they give. Companies who want to join the benefits should enter into a contract with the faculty and allocates funds for scholarships in the form of a donation to support talented students. Students who want to participate in the program must sign an agreement with the faculty including non-disclosure agreement. The model was designed to allow students and companies to work together. First it takes the information provided by the companies related to their industry needs and requirements; and then, based on student's academic information, skills and preferences, connects the companies needs with students who have these specific abilities, knowledge and skills, allowing students to apply and a company to select the best candidates to develop their projects. Some of the problems in this area are described in [5].

Skills are utilized for project recommendation but it is also possible to aggregate them and show summary statistics such as which skills posses average student of our faculty (see Figure 5).

Fig. 5. Average student of FIT CVUT.

Or how skills of students are distributed (Figure 6). Individual skills can be

Fig. 6. Industrial partner can observe distribution of students skills and target better.

O schopnosti - Databáze

Vámi určená úrověň schopnosti je ★★★★☆

Související předměty

Seznam předmětů, které ovlivňují schopnost Databáze. Výsledná úroveň schopnosti se počítá pomocí výsledku v daném předmětu. Jakým poměrem daný předmět ovlivňuje schopnost je vyjádřeno sloupečkem vliv.

Kód	Název	Vliv	Výsledek
MI-BPR	Bezpečnost a bezpečné programování	1.0	
BI-AWD	Administrace webového a DB serveru	1.0	
MI-PDD	Předzpracování dat	1.0	
BI-PCS	Jazyk C# - přístup k datům	1.0	
BI-VWM		1.0	
MI-PDB	Pokročilé databázové systémy	1.0	
BI-AND	Programování pro operační systém Android	1.0	
BI-DBS	Databázové systémy	1.0	
BI-PJV	Programování v Javě	1.0	
BI-EJA	Enterprise java	1.0	
BI-APJ	Aplikační Programování v Javě	1.0	
MI-DSP	Databázové systémy v praxi	1.0	
MI-KDD	Dobývání znalostí z databází	1.0	
BI-SQL	Jazyk SQL	1.0	
MI-VMW	Vyhledávání multimediálního obsahu na webu	1.0	

Rozložení schopnosti u studentů

Graf ukazuje počet studentů k jednotlivým úrovním schopnosti.

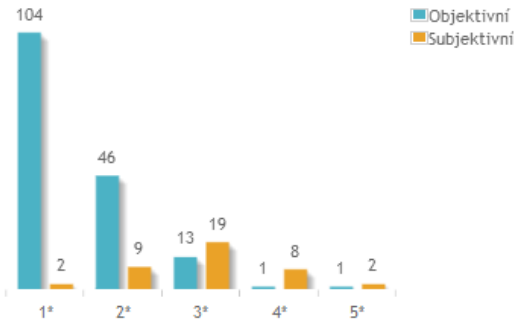

Fig. 7. Skill Databaze is computed from several subjects. The histogram of values is also available both for objective (computed) skills and for subjective (inserted by student).

inspected in the Portal (Figure 7).

5 Data warehouse reports

Fig. 8. Reports can be easily constructed utilizing data warehouse.

There are many more applications where educational data can be utilized once the data warehouse is deployed. For example easy to design and configure reports can signal potential problems before they grow too big [8].

6 Experiences and future improvements

Our Portal runs several months already in the pilot stage. Without advertisement there are already almost fifty companies registered and several assignments completed. We did not expect that bigger problem is motivate students to work on assignments than companies to produce them. We are involving more academics to improve the ontology of skills. We also found out that the architecture of the data warehouse needs to be more robust to changes so we designed and

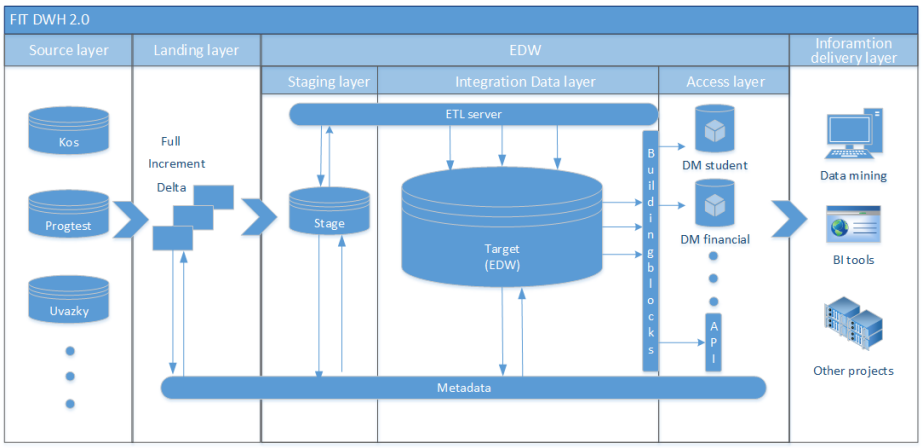

Fig. 9. New design of data warehouse for educational data, that is more robust to changes and easier to extend.

started to implement new version (Figure 9) inspired by [7]. There are several tools that allows rapid prototyping of reports from data warehouse. We work on these reports to present data in a way leading to more informed decisions. There are many more applications where educational data can be utilized once the data warehouse is deployed. We are building a Business Intelligence (BI) infrastructure to be able to report actual development of key performance indicators. Using the data integrated in the warhouse, we can also predict the future behavior, like how many percent of student will pass to the next semester and etc. BI will help us to uncover the behaviour our system in immediate future. Figure 8 demonstrate some of the basic operational dashboards used in our system.

7 Conclusions

Education data can be integrated into data warehouse and subsequently used for several useful purposes - analytic reports are just one of them. In this paper we show how skills can be computed from evaluations and used to recommend relevant students for industrial projects. Understanding and utilizing our data is helping us to increase efficiency and innovate in the right place.

8 Acknowledgments

The development of portal and related technologies was funded by CRP 2012 grant (MSMT) and by the Faculty of information technology, Czech Technical University in Prague. Thanks to Adriana Ramírez for joining our team and documenting portal for student collaboration with industry, Eliška Hrubá for elaborating dashboards generated from data warehouse and Aleš Fišer for designing and implementing reports used in the Portal.

References

1. R. Baker. Data mining. In P. Peterson, E. Baker, and B. McGaw, editors, *International Encyclopedia of Education (Third Edition)*, pages 112 – 118. Elsevier, Oxford, third edition edition, 2010.
2. K. Börner, M. Conlon, J. Corson-Rikert, and Y. Ding. VIVO: A Semantic Approach to Scholarly Networking and Discovery. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 2(1):1–178, Oct. 2012.
3. G. E. etal. Making industry-university partnerships work, lessons from successful collaborations. *2012 Science—Business Innovati on Board AISBL*, 1:1–50, 2012.
4. R. Kimball and M. Ross. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. Wiley, 2013.
5. J. Mnch, F. Fagerholm, P. Johnson, J. Pirttilahti, J. Torkkel, and J. Jarvinen. Creating minimum viable products in industry-academia collaborations. *Lean Enterprise Software and Systems*, 167:137–151, 2013.
6. T. J. T. F. on Computing Curricula Association for Computing Machinery IEEE-Computer Society. Computer science curricula 2013. Technical report, ACM IEEE Computer Society, 2013.
7. P. Ponniah. *Data Warehousing Fundamentals for IT Professionals*. Wiley, 2011.
8. O. Yigitbasioglu and O. Velcu. A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, 13(1):41 – 59, 2012.