

A Comparative Framework to Evaluate Recommender Systems in Technology Enhanced Learning: a Case Study

Matteo Lombardi and Alessandro Marani

School of Information and Communication Technology,
Griffith University,
170 Kessels Road, Nathan, QLD, 4111 Australia
{matteo.lombardi,alessandro.marani}@griffithuni.edu.au

Abstract. When proposing a novel recommender system, one difficult part is its evaluation. Especially in Technology Enhanced Learning (TEL), this phase is critical because those systems influence students or educators in educational tasks. Our research aims to propose a framework for conducting comparative experiments of different recommender systems in a same educational context. The framework is expected to provide the accuracy of subject systems within a single experiment, depicting the benefits of a novel system against others. We also present an application of such framework for a comparative experiment of popular systems in TEL like Google, Slideshare, Youtube, MERLOT, Connexions and ARIADNE. Our results show that the proposed framework has been effective in comparing the accuracy of those systems, with a clear picture of their performance compared one another. Moreover, the results of the experiment can be used as a benchmark when evaluating novel recommender systems in TEL.

Keywords: Recommender Systems, Technology Enhanced Learning, Comparative Evaluation, Evaluation Experiment, Accuracy Performance Evaluation

1 Introduction

Recommender systems are widely studied and developed in literature to assist users in the retrieval of relevant goods and services. Those systems use different approaches to perform the recommendation, mostly content-based filtering, collaborative filtering, knowledge-based filtering and their hybridisations [29]. Such approaches are currently applied for the recommendation of items in several areas including education with Technology Enhanced Learning (TEL) systems [7].

Indeed, many resources available on the Web may represent a potential educational content, therefore educators consider Internet as a place where they can search for materials [15]. However, searching for educational resources is a more complex task than for other goods/services, even for recommender systems

that meet more challenges to perform proper suggestions. The recommendation of educational resources is more complex for i) the many aspects involved in the selection of a resource, and ii) the low amount of educational data available compared to other areas [29]. In addition, users' interests and preferences alone are not enough for a proper recommendation of educational content, because the recommended resources shall meet users' educational context as well as users' characteristics [3].

In general, the evaluation of recommender systems is a critical step when proposing a novel recommendation algorithm, because of i) the many properties that should be evaluated (e.g. accuracy, robustness, scalability), ii) the datasets and iii) the number of users that such evaluations may require [23]. When evaluating a recommender system, three experimental settings are expected: offline experiment, user studies and online experiment [23]. According to Shani and Gunawardana (2011), the former is towards the calibration of the recommendation algorithm and process, or even for comparing the performances of different algorithms. For such phase, a number of existing real datasets or synthetic datasets may be used. The second experiment is towards the analysis of the system behavior with a set of test users. For this experiment, the test users interact with the system performing a set of tasks. In such scenario, the researchers record some quantitative data about the system, the users and the interaction system-user in order to have a first understanding of the system performances and influence on users. Finally, online experiment is for the evaluation of the system in real applications, where it is mainly measured how much the recommendations proposed by the system actually influence the users, as well as other properties. When evaluating recommender systems in TEL, the same three experiments are expected [7].

As the next section reports, there are many contributions about recommender systems in TEL, but there is not a comparison of their performance in similar educational contexts, which makes a comprehensive understanding of their effectiveness difficult. Even if those systems perform well on recommending educational resources considering some specific aspects, it is worth to see whether or not their overall performance is better than current practices. What we find in this study is that generic IR systems such as Slideshare, Google and Youtube are very popular among users of TEL systems and they do not perform bad either (see Section 6). Therefore, it is important to prove that the performance of novel systems are at least the same even if their recommendations are more appropriate for some educational aspects (e.g. educational profile of users). Hence, it would be more interesting to see different recommendation systems compared in a same framework or evaluation setting. In this way, the resulting performance of a novel system can be compared against a set of systems that are quite popular, providing a more clear evidence of the effective progress in the field of recommender systems in TEL. Indeed, in TEL we can find different ways to recommend educational resources [7], thus, for a more effective and expressive evaluation, it is important to use the same experimental setting. In this regard, we suggest a comparative framework which consists of a set of information that

depicts the educational context on which to conduct the comparison of recommender systems in TEL.

Finally, we present an application of our framework for organising a comparative experiment. The objective of such experiment is the evaluation of the accuracy performance of Information Retrieval (IR) systems practically used by users of TEL systems, namely students and educators. The results here reported can be a first starting point for measuring the performance of novel TEL recommender systems. During this contribution we explore the accuracy performance of Google, Youtube and Slideshare, which are very popular among educators and students [4], although they are generic search engines, not focused only on educational resources. In addition to them, we evaluate Connexions¹, MERLOT² and ARIADNE³ which are the most popular Learning Object repositories in literature and practical life of students and educators [4, 14, 25].

2 Background

As said in the introduction, the retrieval of educational resources is a complex task mostly because of the many aspects that are involved in the recommendation process. The literature proposes the Learning Object for the description of educational characteristics of digital content through metadata [26]. Such information is helpful for recommender systems to find a match between the user query and metadata of digital resources. Learning Objects are available on the Web in Learning Object Repositories such as MERLOT, Connexions and ARIADNE. Those systems differently gather and manage Learning Objects that they host. In particular, MERLOT does not actually store the materials that it offers, but it only records the link to the website of the resource and the metadata, as stated on its website⁴. Moreover, it is particularly appreciated for the peer review of its materials. MERLOT currently refers to 40.000 resources. In opposite to MERLOT, Connexions stores the materials, presently over 26.000 Learning Objects according to the information reported on its website⁵. Differently, ARIADNE [10] is a federated repository based on GLOBE⁶, a network of Learning Objects coming from different repositories [19]. So, ARIADNE actually uses Learning Objects available in other repositories but stored with the IEEE LOM standard⁷, which presently is the most popular standard of the metadata of Learning Object [19].

The main contribution of Learning Object Repositories is that the user can search for Learning Objects typing a phrase containing the keywords, then the

¹ <https://cnx.org>

² <http://www.merlot.org>

³ <http://www.ariadne-eu.org>

⁴ http://info.merlot.org/merlothelp/index.htm#merlot_collection.htm accessed on 12/05/2015

⁵ <http://cnx.org/contents> accessed on 12/05/2015

⁶ <http://globe-info.org/>

⁷ IEEE 1484.12.1-2002, IEEE standard for learning object metadata

Learning Object Repository will present all the available resources matching with the query, similarly to Google. The search engines of Learning Object Repositories are more appropriate for recommender systems in TEL, but they can access to a more limited number of resources than Google, Youtube or other big IR systems.

Currently, interesting systems based on Learning Object are proposed with the aim to produce proper recommendations for students and educators [29, 7], some of them also based on user profiling [13, 12, 22]. Although we have those systems, users still mostly rely on generic search engines. In this situation, we believe that a comparative evaluation against generic IR systems represents a stronger evidence of the effectiveness of novel systems. In this regard, a comparative experiment based on a same benchmark provides the research community, industries and end users with a valid motivation to change their attitude towards novel systems and practices.

Table 1. A review of some recommender systems in TEL focused on the evaluation experiments.

System	Online Evaluation	Offline Evaluation	Comparative Evaluation	Accuracy Measures	Predictive Accuracy
[18] (2007)		x			x
[16] (2008) x (50 students)					x
[25] (2010)		x			x
[30] (2011) x (10 students)					x
[2] (2011)		x		x	x
[5] (2012) x (Java course)					x
[21] (2013)		x		x	
[31] (2013) x (24 students)				x	
[9] (2014)		x		x	
[1] (2014)		x		x	x
[22] (2015)		x			x

In literature, we found some surveys of recommender systems in TEL [29, 7, 17] where it is depicted a very good overview of those systems and their approaches. Unfortunately, a comparative evaluation of the resulting performance of such systems is difficult to conduct because of the different experimental settings conducted in those studies. However, the performance of those systems (e.g. accuracy) can still be compared if they are evaluated in the same experiment. To the best of our knowledge, no works either conduct a comparative evaluation of accuracy performance or propose a comparative framework of recommender systems in TEL. In this regard, Table 1 reports some recommender systems in TEL and their experimental settings. As we can see, none of them has conducted a comparative evaluation with other systems, neither with current practices of their end users. In addition, the performance analysis is mostly based on predic-

tive accuracy measures, that however cannot be used for big IR systems for the absence of ratings of their items from the educational point of view. Therefore, in this work we adopt Top-N accuracy measures of recommender systems [8], which are largely used also for the evaluation of generic search engines [28, 11].

3 The proposed comparative framework

The framework proposed in this contribution consists of a set of features used for conducting a comparative experiment of accuracy performance of recommender systems in TEL. In practice, our framework expresses the keywords for describing some educational features that the retrieved resources are expected to meet. The framework must be as generic as possible in order to make the comparison possible for most of the systems; the more the framework is generic, the more IR systems can be evaluated with it. Moreover, the framework must be based on information that is user independent. As result, the framework proposed in

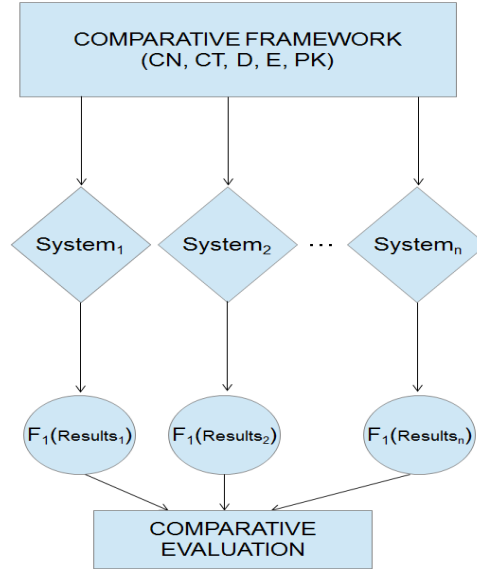


Fig. 1. Structure of a comparative experiment based on the proposed Comparative Framework.

this contribution consists of:

1. Concept Name (CN)
2. Course Title (CT)
3. Difficulty (D)
4. Education Level (E)

5. Prerequisite Knowledge (PK)

Those elements represent a piece of information that is essential when searching for educational resources and independent of user characteristics [24, 12]. Moreover, the order as they are here presented is important too. In fact, the framework submits 5 queries to the subject systems, where in each query is added one of those elements following that order (e.g. Query1 has only CN, then Query2 has CN and CT, and so forth). With this approach we expect to provide a more wide evaluation of the systems with different queries of increasing depth. So, we can have a spectrum of the systems' performance for the different specificity of the queries.

Figure 1 shows the structure of a comparative experiment using the proposed framework. As we can see, an instance of the framework is used to query different systems. Then, it is possible to compare the results using well-known accuracy measures of recommender systems. The framework is proposed with accuracy measures, instead of predictive accuracy or other measures, because it is not easy to find educational datasets for recommender systems in TEL, even more with ratings from users [29, 3]. Being the goal of the comparative experiment the analysis of accuracy performance of subject systems, *Precision* (P) and relative *Recall* (R) [23] have been used as accuracy measures. For an expressive comparative experiment, we suggest to combine P and R in one score using the F_1 score that is, in essence, the harmonic mean of P and R [8, 23]. The formula of the F_1 score is the following:

$$F_1 = \frac{2PR}{P + R} \quad (1)$$

In the next section we provide all the details about the framework and its application for comparing some systems used to retrieve educational material.

4 The framework in action: a case study

In this section we apply the proposed framework to conduct a comparative experiment of accuracy performance of Google, Youtube, Slideshare, Connexions, MERLOT and ARIADNE for the retrieval of resources appropriate for an educational context, which is described by an instance of the framework. We have selected those systems because they are either very popular among users of recommender systems in TEL, such as Google, Youtube, Slideshare and MERLOT [4], or with high attention in literature, like Connexions and ARIADNE [14, 6, 20]. The instance of the framework involved for this evaluation is about a *Java Programming* course for *Undergraduate* students and *Beginners* as difficulty level. All the queries submitted to the subject systems are about the retrieval of material for the concept *Operators* that has *Variables* as prerequisite in such course (derived from the official Java tutorial⁸). Therefore, the instance of the framework for this experiment consists of the following data:

⁸ <https://docs.oracle.com/javase/tutorial/java/index.html>

1. Concept Name: Operators
2. Course Title: Java Programming
3. Difficulty: Beginners
4. Education Level: Undergraduate
5. Prerequisite Knowledge: Variables

From this set of information, we query each subject system five times, each time increasing the depth of the query adding a piece of information of the framework. So, we start with a query formed by only *Operators*, then the second query consists of *Operators* and *Java Programming*, and so on. The last query contains all the elements of the framework. The set of queries used for this comparative experiment is reported in Table 2.

In order to reduce every possible bias due to the automated user profiling that most of the analysed systems perform, we have decided to i) completely delete the cache, history, cookies and any other browsing data of the browser before to perform each query, and ii) execute the queries browsing in private (incognito). The browser used in this experiment is Internet Explorer.

Table 2. Set of queries used to evaluate the performance of the test systems.

Query #	Query	Contextual Information
1	Operators	CN
2	Operators Java programming	CN, CT
3	Operators Java programming beginners	CN, CT, D
4	Operators Java programming beginners undergraduate	CN, CT, D, E
5	Operators Java programming beginners undergraduate variables	CN, CT, D, E, PK

Specifically for this experiment, the same authors (with expertise in teaching Java Programming courses for Undergraduate) decided the relevance of the retrieved items, because of their expertise in Java Programming and their impartiality in evaluating the subject systems (the authors are not linked in any way to any of the analysed systems), as performed in other literature for the evaluation of big IR systems [28].

In addition, only the items displayed in the first page of the results are considered, given the fact that just few users go to the second page or beyond when browsing the results of large IR systems [27]. Given that all the test systems present at least ten results for each page, P and R are calculated according to the relevance of the first ten results (top-10 accuracy measure). As said, the relevance of a retrieved document is established by the same authors with a binary value (1 if relevant, 0 otherwise). A retrieved item is considered relevant when i) it covers all the concept *Operators* and not just a part of it, even if it covers also

other topics, and ii) it complies with all the features of the educational context as described by the framework.

For the computation of the F_1 scores, the set of relevant items of each subject system consists of the union of relevant resources retrieved by the individual system during the execution of the five queries. In the following section, we present the subject systems that are evaluated in our experiment, with a particular focus on how they perform the recommendations.

4.1 Brief overview of subject systems

As said earlier, during the comparative experiment presented in this paper, Google, Youtube, Slideshare, Connexions, MERLOT and ARIADNE are evaluated. With more focus on the data used for their recommendation processes and the structure of such processes, companies like Google Inc. (owner of Google search engine and Youtube platform) and LinkedIn (which is behind Slideshare) do not use only the keywords in the string query to perform their recommendations. Information about the user is deducted even if the person is not logged in, via an automated user profiling.

For example, as stated in the privacy policy of Google Inc.⁹, Google can depict a user profile through the previous queries, visited web pages, Youtube videos watched in the past, the location of the user via the IP address, the device of the user and even the movements of the mouse pointer. This is applied not only for the Google search engine, but also for Youtube. Hence, recommendations are personalised using such user data collected in background of user's activities on the web.

According to the privacy policy of Slideshare¹⁰, also this system collects information about the chronology of visualised content and queries performed by the user. In addition, if the user is logged in the system, Slideshare can deduct different data in several ways. For instance, the sex of users can be acquired analysing their name, and the users' qualifications are used to deduct age, fields of interest and range of income.

Those three subject systems mostly use generic user profiles to personalise the query results, but they do not build or use an educational profile of users. Although those systems are not focused in education, the experimental setting here proposed can still be applied. For the proposed experimentation, educational features of the retrieved items are not required for comparing the relevance of the results, but the comparative framework itself is used for querying the subject systems and then evaluate the relevance of their results; an approach valid also for Google, Slideshare and Youtube. In this manner, the comparative evaluation can show the performance of the subject systems in retrieving educational resources for the same context as described by the instance of the comparative framework.

⁹ <http://www.google.com/intl/en/policies/privacy/>

¹⁰ <https://www.linkedin.com/legal/privacy-policy>

5 Performance analysis

In this section we present and discuss the performance of the subject systems involved in the comparative experiment presented in this contribution. Along the next subsections, we report the results and discussions for each analysed system.

5.1 Google

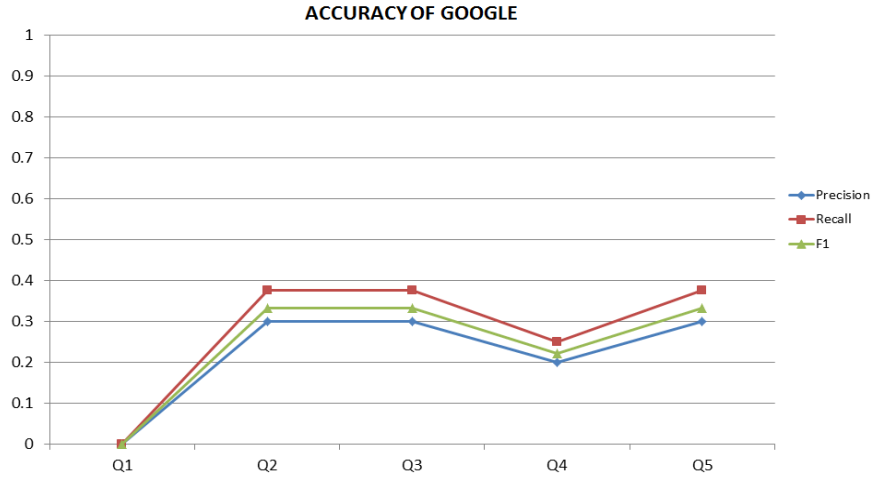


Fig. 2. Results of accuracy performance of Google using the framework.

The chart of the performance of Google search engine is shown in Figure 2. Google has retrieved 8 relevant items after the five queries. As we can see, there is an increment of Precision and Recall until the second query, reaching the best values which are 0.3 and 0.375 respectively. The performance remains steady with query number three, but it presents a drop when the query is expanded with *Education Level*. Adding the *Prerequisite Knowledge* to the query, Precision and Recall rise again to their maximum value. Overall, Google has a F_1 score which is always just over 0.3 throughout the experiment, except for query Q1 and Q4. It is not an excellent performance, but we have to consider that Google dataset is the entire web, not only educational resources, so it is acceptable that it does not have the best performance.

5.2 Slideshare

Throughout the experiment, Slideshare retrieved 5 relevant resources. Figure 3 shows that Precision and Recall of Slideshare have an increasing trend until

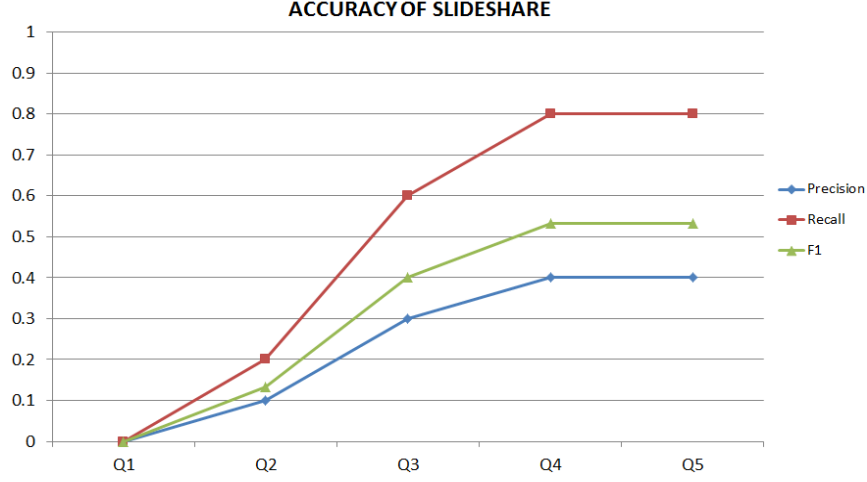


Fig. 3. Results of accuracy performance of Slideshare using the framework.

the fourth query reaching the maximum value of 0.4 for Precision and 0.8 for Recall. The same performance is recorded for the last query. So, in opposite to Google, Slideshare benefits of the *Education Level* element and, overall, it performs better than Google. However, we expected still better results because Slideshare is more focused in education than Google.

5.3 Youtube

The third system involved in this experimentation is Youtube, with 6 relevant items retrieved over the entire experiment, and Figure 4 reports its accuracy performance. We can say that Youtube has a steady trend during the experiment except for the first and the last queries where it is registered a relevant drop of the accuracy. In particular, the first query produces a F_1 score of 0, because the retrieved items were not sufficient to cover the entire concept, but just a part of it. From the second to the fourth query, there is the best performance of Youtube, with Precision=0.2 and Recall just over 0.3. Then, with the last query, Precision falls to 0.1 and Recall to less than 0.2. We can say that Youtube did not perform well against the other systems analysed so far. The main problem of such a low performance of Youtube is due to the granularity of its items, which might be too much fine for educational purposes according to the results of this experiment. In fact, although *Operators* is not a wide concept, most of the retrieved items are short videos which explain only a portion of it.

5.4 Connexions

Throughout the entire experiment Connexions has retrieved 2 relevant items and the accuracy performance is shown in Figure 5. For this system, we can see that

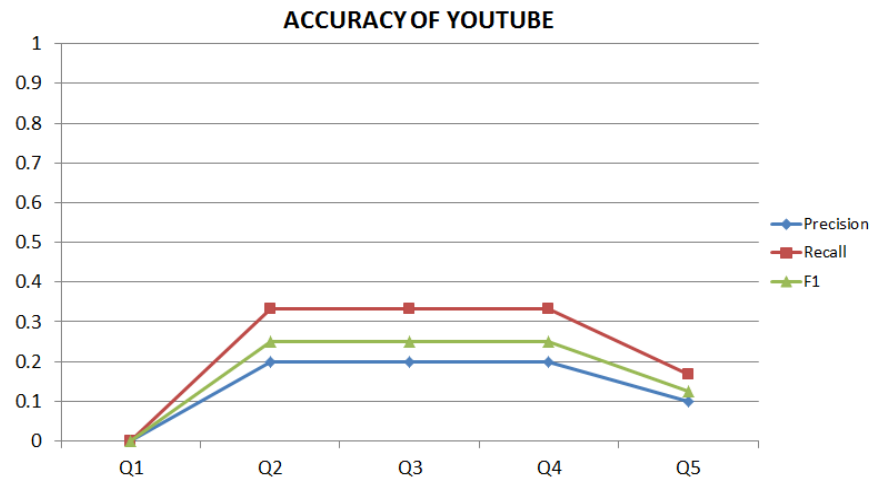


Fig. 4. Results of accuracy performance of Youtube using the framework.

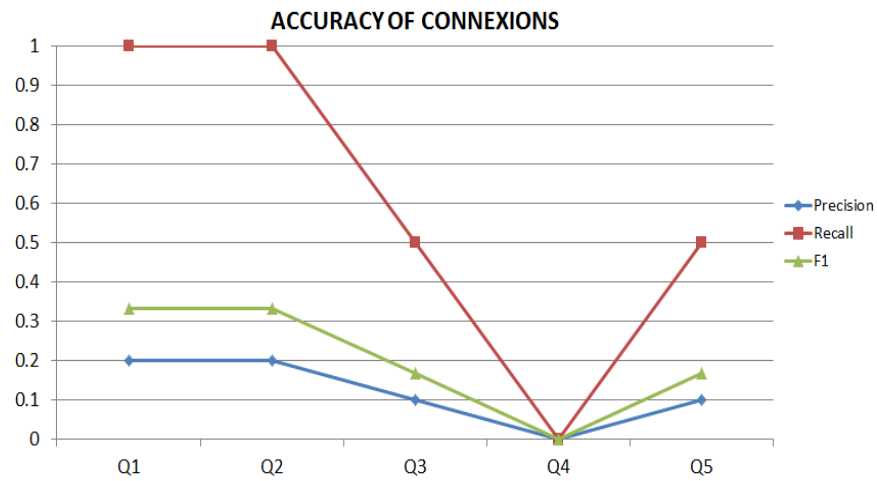


Fig. 5. Results of accuracy performance of Connexions using the framework.

the best values for precision and recall are achieved straight away with the first two queries, namely Recall=1 and Precision=0.2, because the only two relevant items are both retrieved in these two first queries. After that, the trend starts to fall until 0 when, in query Q4, the *Education Level* is added. A significant increment is recorded for the fifth query. The main problem of this system is the low availability of relevant material. In fact, the highest recorded value of recall is 1 because the two relevant items have been retrieved right away in the first two queries. However, it is interesting to note that it is the unique system in this experiment that reached its maximum accuracy in the first query.

5.5 MERLOT and ARIADNE

Both MERLOT and ARIADNE had issues in the retrieval of Learning Objects during this experiment. In particular, MERLOT retrieved items only for the first query, but none of them were relevant. No results have been reported from the second query to the last one. Similarly, ARIADNE was not able to retrieve any resource for any query, and, in addition, it mostly retrieved broken links. For these reasons, we could not produce the charts about the performance of MERLOT and ARIADNE.

6 Summarising the results: comparative analysis

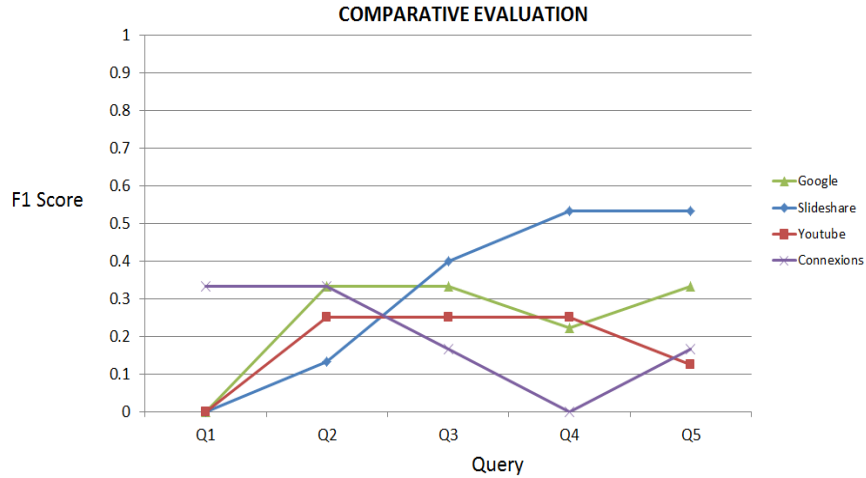


Fig. 6. Comparative evaluation of Google, Slideshare, Youtube and Connexions based on F_1 scores resulting from the application of an instance of the comparative framework.

During our analysis we have been able to retrieve relevant items from the majority of the tested IR systems, but with different performance for each query.

From the charts presented in the previous section, we can see that Slideshare achieved the best results in terms of precision and recall, especially when nearly all the elements of the framework were in the query. With the application of the comparative framework here proposed, we can have a very significant and expressive performance of all the four systems of this experiment, as Figure 6 clearly shows. In fact, just looking on that figure we can conclude that Connexions and Google have the best accuracy performance for low-specific queries (Q1 and Q2). From the third query on, it is evident that Slideshare has a remarkable and overwhelming performance compared to the other three systems, with an increasing performance as the queries are more specific.

In conclusion of this experiment, we have shown the benefit of the proposed framework for comparing recommender systems in TEL. Such approach is simple but very effective in properly conducting comparative experiments and producing comparable results. Moreover, the results found in this experiment can be a benchmark for comparative evaluation of novel systems in TEL. Indeed, new recommender systems in TEL are expected to have an overall accuracy performance higher than 0.53 in terms of F_1 score. That value is the highest F_1 score observed in this experiment, achieved by Slideshare with the fourth and fifth query (refer to Figure 6). Finally, the comparative framework can also help to understand what educational information enhance the accuracy performance of a novel system.

6.1 Limitations of the experiment

The analysis presented in this contribution has been performed on several IR systems with the most possible attention to avoid any bias or other condition that could invalidate the experiment, but it is limited to only one domain. However, this limitation does not weak this experiment, because Java programming courses are taught in several universities around the world in Undergraduate classes, so the instance of the framework used in this experiment is a very popular scenario in the real world. Furthermore, it is highly probable that an educator or student could use the discussed IR systems to look for educational resources about the topic included in our analysis. More important, in this paper we have been able to show how it is practically possible to conduct a comparative experiment using the framework here proposed, and the effectiveness of the framework itself in comparing the performance of different systems.

7 Conclusions and future work

In this contribution we have seen the importance and feasibility of conducting a comparative experiment of recommender systems in TEL, towards a better understanding of their benefits for users. As result of the review of the literature, no comparative studies of such systems have been conducted in the field of TEL. Moreover, it is evident the lack of comparative evaluations of new systems against current solutions. Such evaluations are important when presenting

a novel system in order to provide researchers, industries and even the end users with a clear improvement of the recommendation service provided by the novel system. The framework proposed in this contribution provides the stakeholders of recommender systems in TEL with a tool for comparing the accuracy performance of new systems against what they currently use. Even for researchers, this framework is fundamental for conducting comparative experiments for better insights of the performance of their work, which is the main goal of this paper.

In addition, an application of the framework in a popular case study is presented, namely the recommendation of educational resources for a Java Programming course (refer to Section 4 for more details). Using our framework, a comparative experiment of Google, Slideshare, Youtube, Connexions, MERLOT and ARIADNE has been conducted. Figure 6 summarises the results of the comparative experiment based on our framework, where those systems have been compared one another in a single experiment and in the same educational context. In that figure, MERLOT and ARIADNE are not reported because of their problems in performing the steps of the experiment (refer to Section 5.5 for more details).

Another important finding of this paper is the result of such experiment, which is a very significant benchmark when conducting the evaluation of future novel recommender systems in TEL. In fact, researchers can repeat the comparative experiment with the case study here analysed, and use the results of this experiment for demonstrating the actual contribution of their systems.

In conclusion, with the application of the proposed comparative framework, a comparative experiment has been successfully conducted for the comparison of accuracy performance of systems with different characteristics. We expect other applications of the framework for a better understanding of its effectiveness for the evaluation of other systems.

References

1. Deepa Anand and Bonson Sebastian Mampilli. Folksonomy-based fuzzy user profiling for improved recommendations. *Expert Systems with Applications*, 41(5):2424–2436, 2014.
2. Jesus Bobadilla, Fernando Ortega, Antonio Hernando, and Javier Alcalá. Improving collaborative filtering recommender system results and performance using genetic algorithms. *Knowledge-based systems*, 24(8):1310–1316, 2011.
3. Jorge Bozo, Rosa Alarcón, and Sebastian Iribarra. Recommending learning objects according to a teachers? contex model. In *Sustaining TEL: From innovation to learning and practice*, pages 470–475. Springer, 2010.
4. Isabelle Brent, Graham R Gibbs, and Anna Katarzyna Gruszczynska. Obstacles to creating and finding open educational resources: the case of research methods in the social sciences. *Journal of Interactive Media in Education*, 2012(1):Art–5, 2012.
5. Ana Casali, Valeria Gerling, Claudia Deco, and Cristina Bender. A recommender system for learning objects personalized retrieval. *Educational Recommender Sys-*

- tems and Technologies: Practices and Challenges*, Hershey, PA: Information Science Reference, pages 182–210, 2012.
6. Cristian Cechinel and Xavier Ochoa. A brief overview of quality inside learning object repositories. In *Proceedings of the XV International Conference on Human Computer Interaction*, page 83. ACM, 2014.
 7. Hendrik Drachsler, Katrien Verbert, Olga Santos, and Nikos Manouselis. Panorama of recommender systems to support learning. *status: accepted*, 2015.
 8. Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1):5–53, 2004.
 9. Pythagoras Karampiperis, Antonis Koukourikos, and Giannis Stoitsis. Collaborative filtering recommendation of educational content in social environments utilizing sentiment analysis techniques. In *Recommender Systems for Technology Enhanced Learning*, pages 3–23. Springer, 2014.
 10. Joris Klerkx, Bram Vandeputte, Gonzalo Parra, José Luis Santos, Frans Van Assche, and Erik Duval. How to share and reuse learning resources: the ariadne experience. In *Sustaining TEL: From Innovation to Learning and Practice*, pages 183–196. Springer, 2010.
 11. Chunyan Liang. User profile for personalized web search. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on*, volume 3, pages 1847–1850. IEEE, 2011.
 12. Carla Limongelli, Matteo Lombardi, Alessandro Marani, and Filippo Sciarrone. A teacher model to speed up the process of building courses. In Masaaki Kurosu, editor, *Human-Computer Interaction. Applications and Services*, volume 8005 of *Lecture Notes in Computer Science*, pages 434–443. Springer Berlin Heidelberg, 2013.
 13. Carla Limongelli, Matteo Lombardi, Alessandro Marani, and Filippo Sciarrone. A teaching-style based social network for didactic building and sharing. In H.Chad Lane, Kalina Yacef, Jack Mostow, and Philip Pavlik, editors, *Artificial Intelligence in Education*, volume 7926 of *Lecture Notes in Computer Science*, pages 774–777. Springer Berlin Heidelberg, 2013.
 14. Carla Limongelli, Alfonso Miola, Filippo Sciarrone, and Marco Temperini. Supporting teachers to retrieve and select learning objects for personalized courses in the moodle_ls environment. In *Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on*, pages 518–520. IEEE, 2012.
 15. Stephen Maloney, Alan Moss, Jennifer Keating, George Kotsanas, and Prue Morgan. Sharing teaching and learning resources: perceptions of a university’s faculty members. *Medical education*, 47(8):811–819, 2013.
 16. Eleni Mangina and John Kilbride. Evaluation of keyphrase extraction algorithm and tiling process for a document/resource recommender within e-learning environments. *Computers & Education*, 50(3):807–820, 2008.
 17. Nikos Manouselis, Hendrik Drachsler, Riina Vuorikari, Hans Hummel, and Rob Koper. Recommender systems in technology enhanced learning. In *Recommender systems handbook*, pages 387–415. Springer, 2011.
 18. Nikos Manouselis, Riina Vuorikari, and Frans Van Assche. Simulated analysis of collaborative filtering for learning object recommendation. In *Proceedings of the 1st Workshop on Social Information Retrieval for Technology Enhanced Learning*, pages 27–35, 2007.
 19. Xavier Ochoa, Joris Klerkx, Bram Vandeputte, and Erik Duval. On the use of learning object metadata: The globe experience. In *Towards ubiquitous learning*, pages 271–284. Springer, 2011.

20. Nikolaos Palavitsinis, Nikos Manouselis, and Salvador Sanchez-Alonso. Metadata quality in learning object repositories: a case study. *Electronic Library, The*, 32(1):62–82, 2014.
21. Mojtaba Salehi. Application of implicit and explicit attribute based collaborative filtering and bidirectional learning for learning resource recommendation. *Data & Knowledge Engineering*, 87:130–145, 2013.
22. Stylianos Sergis and Demetrios Sampson. Learning object recommendations for teachers based on elicited ICT competence profiles. *Learning Technologies, IEEE Transactions on*, 2015.
23. Guy Shani and Asela Gunawardana. Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer, 2011.
24. Lee S Shulman. Those who understand: Knowledge growth in teaching. *Educational researcher*, pages 4–14, 1986.
25. Miguel-Ángel Sicilia, Elena García-Barriocanal, Salvador Sánchez-Alonso, and Cristian Cechinel. Exploring user-based recommender results in large learning object repositories: the case of merlot. *Procedia Computer Science*, 1(2):2859–2864, 2010.
26. Mike Sosteric and Susan Hesemeier. When is a learning object not an object: A first step towards a theory of learning objects. *The International Review of Research in Open and Distance Learning*, 3(2), 2002.
27. Amanda Spink and Bernard J Jansen. Searching multimedia federated content web collections. *Online Information Review*, 30(5):485–495, 2006.
28. Duygu Tumer, Mohammad Ahmed Shah, and Yiltan Bitirim. An empirical evaluation on semantic search performance of keyword-based and semantic search engines: Google, yahoo, msn and hakia. In *Internet Monitoring and Protection, 2009. ICIMP'09. Fourth International Conference on*, pages 51–55. IEEE, 2009.
29. Katrien Verbert, Nikos Manouselis, Xavier Ochoa, Martin Wolpers, Hendrik Drachler, Ivana Bosnic, and Erik Duval. Context-aware recommender systems for learning: a survey and future challenges. *Learning Technologies, IEEE Transactions on*, 5(4):318–335, 2012.
30. Xin Wan and Toshio Okamoto. Utilizing learning process to improve recommender system for group learning support. *Neural Computing and Applications*, 20(5):611–621, 2011.
31. Alfredo Zapata, Víctor H Menéndez, Manuel E Prieto, and Cristóbal Romero. A framework for recommendation in learning object repositories: An example of application in civil engineering. *Advances in Engineering Software*, 56:1–14, 2013.