

Sherlock: a Semi-Automatic Quiz Generation System using Linked Data

Dong Liu¹ and Chenghua Lin²

¹ BBC Future Media & Technology - Knowledge & Learning, Salford M50 2QH, UK,
Dong.Liu@bbc.co.uk

² Department of Computing Science, University of Aberdeen, AB24 3UE, UK
chenghua.lin@abdn.ac.uk

Abstract. This paper presents Sherlock, a semi-automatic quiz generation system for educational purposes. By exploiting semantic and machine learning technologies, Sherlock not only offers a generic framework for domain independent quiz generation, but also provides a mechanism for automatically controlling the difficulty level of the generated quizzes. We evaluate the effectiveness of the system based on three real-world datasets.

Keywords: Quiz Generation, Linked Data, RDF, Educational Games

1 Introduction

Interactive games are effective ways of helping knowledge being transferred between humans and machines. For instance, efforts have been made to unleash the potential of using Linked Data to generate educational quizzes. However, it is observed that the existing approaches [1, 2] share some common limitations that they are either based on domain specific templates or the creation of quiz templates heavily relies on ontologist and Linked Data experts. There is no mechanism provided to end-users to engage with customised quiz authoring.

Moreover, a system that can generate quizzes with different difficulty levels will better serve users' needs. However, such an important feature is rarely offered by the existing systems, where most of the practices simply select the distractors (i.e., the wrong candidate answers) at random from an answer pool (e.g., obtained by querying the Linked Data repositories). Some work has attempted to determine the difficulty of a quiz but still it is simply based on assessing the popularity of a RDF resource, without considering the fact that the difficulty level of a quiz is directly affected by semantic relatedness between the correct answer and the distractors [3].

In this paper, we present a novel semi-automatic quiz generation system (Sherlock) empowered by semantic and machine learning technologies. Sherlock is distinguished from existing systems in a few aspects: (1) it offers a generic framework for generating quizzes of multiple domains with minimum human effort; (2) a mechanism is introduced for controlling the difficulty level of the generated quizzes; and (3) an intuitive interface is provided for engaging users

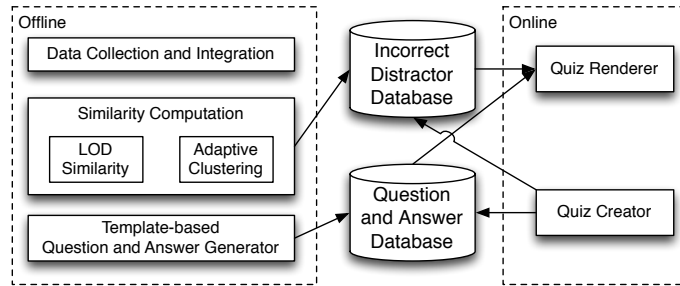


Fig. 1. Overall architecture of Sherlock.

in creating customised quizzes. The live Sherlock system can be accessed from <http://sherlock.pilots.bbconnectedstudio.co.uk/>¹.

2 System Architecture

Fig. 1 depicts an overview of the Sherlock framework, in which the components are logically divided into two groups: online and offline. These components can interact with each other via shared databases which contain the information of the questions, correct answers and distractors (i.e., incorrect answers).

Data Collection and Integration: We collected RDF data published by DBpedia and the BBC. These data play two main roles, i.e., serving as the knowledge base for quiz generation and used for calculating the similarity scores between objects/entities (i.e., answers and distractors).

Similarity Computation: The similarity computation module is the core for controlling the difficulty level of quiz generation. It first accesses the RDF store, and then calculates the similarity scores between each object/entity pairs. In the second step, the module performs K -means clustering to partition the distractors into different difficulty levels according to their Linked Data Semantic Distance (LSDS) [4] scores with the correct answer of a quiz. In the preliminary experiment, we empirically set $K=3$ corresponding to three difficulty levels, i.e. “easy”, “medium” and “difficult”.

Template-based Question and Answer Generator: This module automates the process of generating questions and the correct answers. Fig. 2(a) demonstrates the instantiation of an example template: “Which of the following animals is $\{?animal_name\}$?”, where the variable is replaced with `rdfs:label` of the animal. The generated questions and answers will be saved in the database.

Quiz Renderer: The rendering module first retrieves the question and the correct answer from the database, and then selects suitable distractors from the entities returned by the similarity computation module. Fig. 2(a) shows the module’s intuitive gaming interface, as well as a distinctive feature for tuning up or down the quiz difficulty level dynamically, making Sherlock able to better serve the needs of different user groups (e.g., users of different age and with different

¹ For the best experiences, please use Safari or Opera to access the demo.

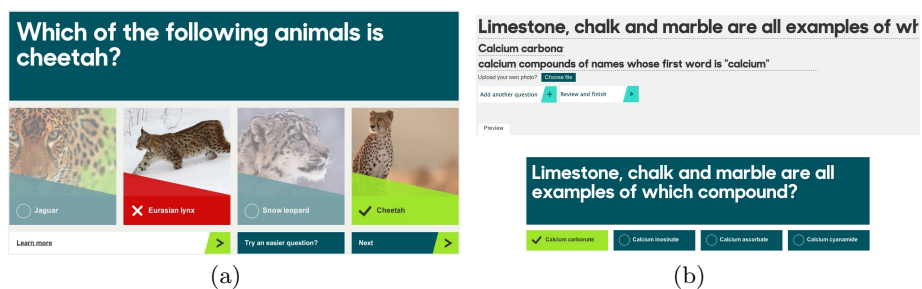


Fig. 2. (a) User interface for playing a quiz; (b) User interface for creating a quiz.

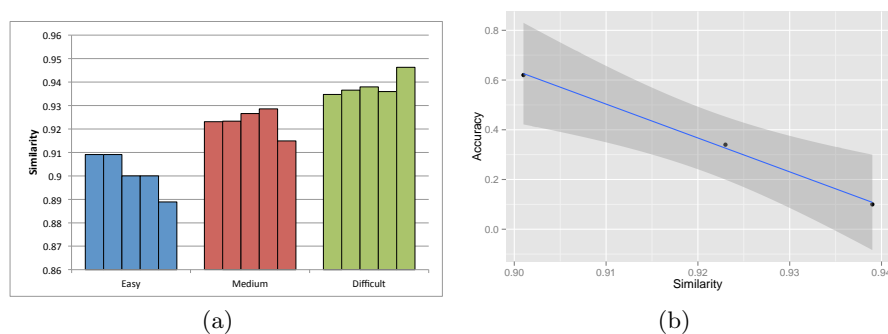


Fig. 3. (a) Averaged similarity of the testing quizzes (Wildlife domain); (b) Correlation measure of the Wildlife domain ($r = -0.97, p < 0.1$).

educational background). Furthermore, to enhance a user’s learning experience, the “*learn more*” link on the bottom left of the interface points to a Web page containing detailed information about the correct answer (e.g., Cheetah).

Quiz Creator: Fig. 2(b) depicts the quiz creator module, which complements the automatic quiz generation by allowing users to create customised quizzes with more diverse topics and to share with others. Quiz authoring involves three simple steps: 1) write a question; 2) set the correct answer (distractors are suggested by the Sherlock system automatically); and 3) preview and submit. For instance, one can take a picture of several ingredients and let people guess what dish one is going to cook. The quiz creator interface can be accessed from <http://sherlock.pilots.bbconnectedstudio.co.uk/#/quiz/create>.

3 Empirical Evaluation

This demo aims to show how Sherlock can effectively generate quizzes of different domains and how well a standard similarity measure can be used to suggest quiz difficulty level that matches human’s perception. The hypothesis is that if some objects/entities have higher degree of semantic relatedness, their differences would be subtle and hence more difficult to be disambiguated, and vice versa.

We investigated the correlation between the difficulty level captured by the similarity measure and that perceived by human. To test our hypothesis, a group of 10 human evaluators were presented with 45 testing quizzes generated by Sherlock based on the BBC Wildlife domain data, i.e., 15 quizzes per difficulty level. Next the averaged pairwise similarity between the correct answer and distractors of each testing quiz were computed, as shown in Fig. 3(a). Fig. 3(b) demonstrates that the quiz test accuracy of human evaluation indeed shows a negative correlation ($r = -0.97$, $p < 0.1$) with the average similarity of the quiz answer choices (i.e., each datapoint is the averaged value over 15 quizzes per difficulty level). This suggests that LDS is an appropriate similarity measure for indicating quiz difficulty level, which inlines with our hypothesis.

In another set of experiments, we evaluated Sherlock as a generic framework for quiz generation, in which the system was tested on structural RDF datasets from three different domains, namely, BBC Wildlife, BBC Food and BBC Your-Paintings², with 321, 991 and 2,315 quizzes automatically generated by the system for each domain respectively. Benefiting from the domain-independent similarity measure (LDS), Sherlock can be easily adapted to generate quizzes of new domains with minimum human efforts, i.e., no need to manually define rules or rewrite SPARQL queries.

4 Conclusion

In this paper, we presented a novel generic framework (Sherlock) for generating educational quizzes using linked data. Compared to existing systems, Sherlock offers a few distinctive features, i.e., it not only provides a generic framework for generating quizzes of multiple domains with minimum human effort, but also introduces a mechanism for controlling the difficulty level of the generated quizzes based on a semantic similarity measure.

Acknowledgements

The research described here is supported by the BBC Connected Studio programme and the award made by the RCUK Digital Economy theme to the dot.rural Digital Economy Hub; award reference EP/G066051/1. The authors would like to thank Ryan Hussey, Tom Cass, James Ruston, Herm Baskerville and Nava Tintarev for their valuable contribution.

References

- [1] Damljanovic, D., Miller, D., O’Sullivan, D.: Learning from quizzes using intelligent learning companions. In: WWW (Companion Volume). (2013) 435–438
- [2] Álvaro, G., Alvaro, J.: A linked data movie quiz: the answers are out there, and so are the questions [blog post]. <http://bit.ly/linkedmovies> (2010)
- [3] Waitelonis, J., Ludwig, N., Knuth, M., Sack, H.: WhoKnows? - evaluating linked data heuristics with a quiz that cleans up dbpedia. *International Journal of Interactive Technology and Smart Education (ITSE)* **8** (2011) 236–248
- [4] Passant, A.: Measuring semantic distance on linking data and using it for resources recommendations. In: AAAI Symposium: Linked Data Meets AI. (2010)

² <http://www.bbc.co.uk/nature/wildlife>, <http://www.bbc.co.uk/food> and <http://www.bbc.co.uk/arts/yourpaintings>