YouUnderstood.me? Personalized Readability-based Retrieval of Online Materials for Students and Educators

Ion Madrazo
Computer Science Department
Boise State University
Boise, Idaho, USA
ionmadrazo@boisestate.edu

Maria Soledad Pera
Computer Science Department
Boise State University
Boise, Idaho, USA
solepera@boisestate.edu

ABSTRACT

K-12 students and educators make use of online resources to satisfy their academic information needs on a daily basis. Unfortunately, they are forced to spend large amounts of time seeking for adequate materials. For students, retrieving materials that are outside their comprehension level can be discouraging. For educators, finding materials for curriculum development that suit their students' individual reading abilities can also be challenging. We present *YouUnderstood.me*, a web application that makes use of information retrieval, natural language processing, and machine learning to provide an enhanced web search environment that promotes the synergy between students and educators and helps them locate materials that match the reading abilities of individual students. *YouUnderstood.me* offers: (1) a search interface that combines a search module, a search intent module, and a readability-based filtering module, which leads to the fast retrieval of documents from different sources, (2) a tracking module that keeps records of students' feedback on retrieved results, in terms of degrees of comprehension accomplished, which in turn informs educators of the change in readability levels of their respective students over time, and (3) a readability analysis module that enables educators on estimate the reading level of snippets of texts. The combination of the described modules make *YouUnderstood.me*, to the best of our knowledge, the first tool specifically designed for educational environments that simultaneously aids both teachers and students dealing with personalized online searches.

CCS Concepts

•**Information systems → Personalization; Search interfaces;** •**Social and professional topics → Children;**

Keywords

Search interfaces; Filtering; Readability; Educational tools

1. INTRODUCTION

The use of Web technologies is increasingly becoming a relevant aspect for children education, both because it enhances the class environment and introduces children, from early stages of their lives, into today's information society [10]. K-12[1] students use the internet on a daily basis to locate materials that can help them with their academic tasks, from finding information for a class presentation to discovering the meaning of a new word. For this purpose, they often turn to search engines and online catalogs to retrieve materials that can satisfy their information needs, including news articles, books, or term definitions. However, even if search engines are successful in terms of retrieving relevant resources that are of interest to a general audience, they can be less so when used by niche users [11], such as K-12 students. Unlike average users, K-12 students have varied reading abilities, a fact that is usually ignored by popular search engines. As a result, students can often get discouraged when they try to read retrieved contents that are outside their comprehension level, whether being too easy or too difficult for them to understand. Therefore, providing them tools to seek for adequate materials they can actually understand is imperative.

Within the academic environment, students are not the only ones facing the problem of locating adequate reading materials that simultaneously match information needs and reading abilities. Educators also face several challenges when looking for materials for their classes, which translates into unnecessary time spent online to accomplish this goal. For example, even in a same grade class, students' reading skills can differ significantly, so not all of them should be given the same texts. As a result, a fair amount of personalization is required which educators are expected to handle in a daily basis. Unfortunately, given the high number of students in each class, this task can become impossible to tackle.

YouUnderstood.me is a web application designed to help both instructors and students in the process of finding online reading materials. The application is centered on the use of readability formulas that together with a search engine make the task of looking for leveled reading materials effective and efficient. By requiring students to log in prior to conducting searches *YouUnderstood.me* can keep track of the feedback they give about the degree of understanding of the read materials (too easy/OK/too complicated). This

---

[1]K-12 refers to the publicly-supported school grades prior to college in the education systems from United States of America, Canada and other countries.
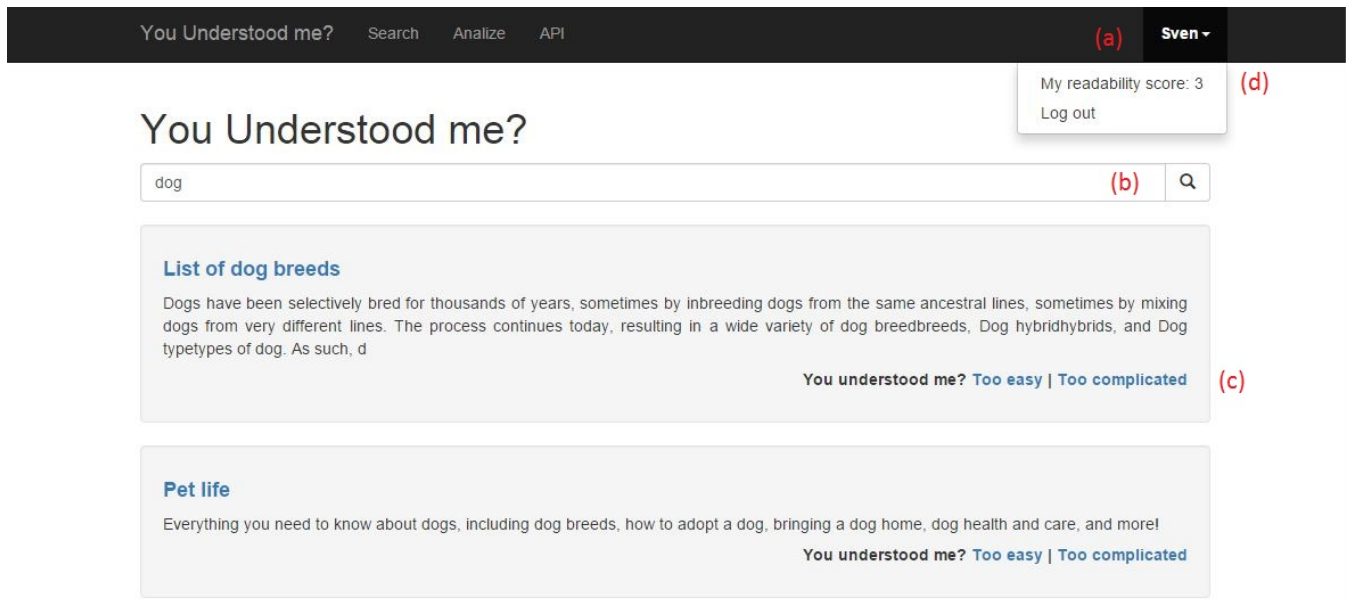
Figure 1: Screenshot of *YouUnderstood.me*'s search interface for students

enables the application to make predictions about the readability score for each student, which can be used by both students and educators for speeding-up the process of searching adequate materials. Furthermore, the application integrates a search interface that allows both stakeholders to seek materials filtered by readability scores from: (i) commercial search engines, such as Google, (ii) public data sources, such as Wikipedia and (iii) local resources, such schools' library catalogs.

Besides the search interface, instructors also have access to an analysis page, where they can submit texts and estimate their readability score, based on a wide range of readability formulas provided within the application. This tool, together with the track of readability scores of students, helps teachers make sure the reading materials they select are adequate for the respective class or an individual student.

The novelty of *YouUnderstood.me* lays on how its different submodules are combined in order to create an enhanced, personalized, search environment that promotes the synergy between students and teachers to help them locate materials that match the reading abilities of each student. To the best of our knowledge, *YouUnderstood.me* is the first application that tackles the issue of reading material retrieval as a whole. Starting from the assessment of an individual student's readability, and ending with the retrieval of adequate materials, all modules of *YouUnderstood.me* work in cooperation to improve the way reading materials are searched.

2. YOUUNDERSTOOD.ME

Regardless of the stakeholder or the task that needs to be performed, *YouUnderstood.me* takes advantage of readability assessment formulas, for determining the level of difficulty of a reading material. Different approaches have been followed in the literature for determining a text's complexity or readability. Most approaches, focus their work in determining the readability of text snippets. Those systems vary from very simple ones [3], which make use of shallow fea-

tures, such as the average number of words per sentence or the average length of terms, to more complex ones [5, 4, 1] , which are mostly based on supervised learning techniques and features extracted using Natural language processing. Those tools, however, have shown to be of little use in contexts where the text of the corresponding material has reduced accessibility, both because the text is not publicly accessible or because it shows a structure not as simple to tackle. Consequently, different works have been done in more specialized contexts such as book [2, 9] or web page retrieval [7], where the systems presented take advantage of domain dependent features.

YouUnderstood.me integrates a readability assessment module than can make use of different metrics and resources simultaneously, aiming to handle a more diverse amount of reading materials. The methods with which *YouUnderstood.me* can predict text complexity are the following:

- **External metrics.** *YouUnderstood.me* is compatible with the most popular readability metrics among the American education system, such as AR[2] or Lexile[3]. The aforementioned metrics are widely used for measuring the readability of books from children curricula. This permits *YouUnderstood.me* retrieve books, which would be difficult to handle otherwise, because of the inability to get access to the contents of copyrighted material.

- **Traditional formulas.** Historically used by teachers for manually determining the readability level of a reading material, traditional formulas such as Flesh [3], Fog[6] and Flesh-Kincaid [3], are also supported by *YouUnderstood.me.*

- **MRAS.** MRAS (Multilingual Readability assessment system)[8] is a state-of-the-art readability assessment

---

[2]www.acceleratelearning.com
[3]www.lexile.com

Figure 2: Screenshot of *YouUnderstood.me*'s search interface for educators

system, which is capable of detecting the input language of a text on-the-fly and providing the corresponding readability score. MRAS is based on a supervised learning paradigm, that makes use of more than a hundred features for readability prediction.

The stakeholders of *YouUnderstood.me* can benefit from the aforementioned readability scores in different ways. On the one side, **students** can make use of a **search engine** (as illustrated in figure 1) which provides several features that are aimed at helping them in the process of seeking adequate reading materials. In order to use it, the student needs to be logged in (a), which permits the application personalize his search experience.

Each query a student formulates is first treated with an ad-hoc search intent module that is capable to solve issues that usually arise while processing children queries and that can lead children to retrieve poor results, such as misspelling, children popular culture terms or too long queries. Processed queries are submitted to a search engine. Currently two methods of search are implemented: one that makes use of the Google Search API and another one based on the Apache Lucene[4] framework.

Documents retrieved by the search engine are analyzed and filtered based on the readability requirements of each individual student. The documents can be both dynamically or statically analyzed, depending on their source and the search tool used. For example, documents are analyzed on the fly when using Google. This requires the readability prediction to be fast, forcing the application to use a lightweight readability formula such as Flesch or FOG. However, when using Lucene, documents can be previously analyzed, enabling the use of more complex readability prediction tools, such as MRAS.

Once the student has read the chosen material, he can provide feedback on it (c), helping the system improve the filtering for future searches. This feedback is given by means of a three option question which permits the student state

---

[4]lucene.apache.org/

whether the reading material was too easy, OK, or too complicated. This information is provided to *YouUnderstood.me*'s tracking module to create predictions about individual users' readability skills.

The **tracking module** is currently based on a trivial method of approximation which simply increases the student's readability score every time he finds a recommended reading material too easy, and decreases it, every time he finds it too complicated. However, we are aware that more precise methods exist and plan to implement them in the near future. An updated prediction of the system regarding each student can be seen on the profile menu (d), allowing each student to keep track of the progress he is making.

Educators can also make use of the **search interface**, which is adapted to their use (see figure 2). The educator view of the search engine enables the user choose the readability level he wants. Moreover, the search engine no longer makes use of the search intent module, given that teachers are supposed to find the correct words for what they are seeking, and adding this extra layer of filtering would only hinder their work. Furthermore, the readability level for filtering the reading materials is no longer decided by the system, giving the instructor the option for choosing the level of challenge he want for his students. The feedback options are no longer available either, since there is no point in evaluating the educator's readability.

Educators can also use the material analysis tool (see figure 4) that allows them to analyze snippets of text. This module allows choosing the readability formula desired for the analysis, among the formulas that have already been mentioned.

Finally, a student tracking tool is provided for educators (fig. 3). This tool provides a fast way to discover the individual and overall reading skills of students in the class. The educator can view a prediction for individual student's reading skills or go deeper and view the feedback that the student gave for each of the reading materials.

3. CONCLUSION AND FUTURE WORK

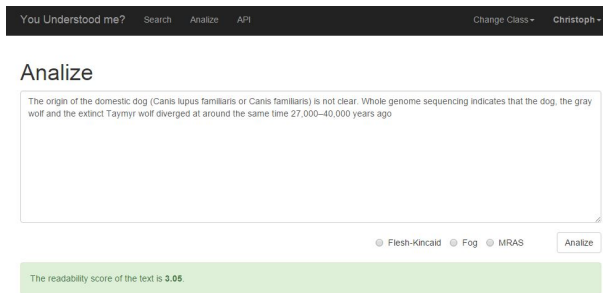**Figure 3: Screenshot of student list for educator**

Figure 4: Screenshot of the analysis page of *YouUndestood.me*

We have introduced *YouUnderstood.me*, a web application that simplifies the task of seeking for adequate online resources in terms of readability levels, within an academic environment. *YouUnderstood.me* adopts a modular design, which facilitates the task of simultaneously integrating different readability prediction strategies, material sources , making it extensible and versatile. It provides an environment tailored to different types of users: students and educators. Students can benefit from a personalized search interface which combines a search intent module and retrieved results filtered by the corresponding readability level of each individual student. At the same time, educators can take advantage of three different tools, i.e., a search interface, which aids educators in locating materials tailored to individual students, a readability analysis tool, which can help them in writing of tests or directions for homework assignments and ensuring that their students can understand them, and a student tracking system, which speeds up the process of locating reading materials suitable to the abilities of each student in class. All the mentioned modules work in cooperation to create, to the best of our knowledge, the first educational tool that tackles the reading material seeking problem as a whole.

On the future, we would like to focus our work on enhancing the synergy between educators and students even more. We would like to give educators a way to validate students' queries, so that students can know which mistakes were included in their queries. This way, students would progressively improve their query formulation skills. We would also want to enhance *YouUnderstood.me* to encourage collaborative searches among students and educators. All this features would lead *YouUnderstood.me* to be an even more versatile tool, a tool that would truly make students and educators understand each other.

4. REFERENCES

[1] F. Dell'Orletta, S. Montemagni, and G. Venturi. Read-it: Assessing readability of italian texts with a view to text simplification. In *ACL Workshop on speech and language processing for assistive technologies*, pages 73–83, 2011.

[2] J. Denning, M. S. Pera, and Y.-K. Ng. A readability level prediction tool for k-12 books. *JASIST*, 2015.

[3] R. Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948.

[4] T. François and C. Fairon. An ai readability formula for french as a foreign language. In *ACL EMNLP-CoNLL*, pages 466–477, 2012.

[5] I. Gonzalez-Dios, M. J. Aranzabe, A. D. de Ilarraza, and H. Salaberri. Simple or complex? assessing the readability of basque texts. In *COLING*, pages 334–344, 2014.

[6] R. Gunning. The technique of clear writing. 1952.

[7] T. P. Lau and I. King. Bilingual web page and site readability assessment. In *WWW*, pages 993–994, 2006.

[8] I. Madrazo and S. Pera. Steps towards multilingual readability assessment. In press.

[9] M. S. Pera and Y.-K. Ng. Automating readers' advisory to make book recommendations for k-12 readers. In *ACM RecSys*, pages 9–16, 2014.

[10] A. Sadaf, T. J. Newby, and P. A. Ertmer. Exploring pre-service teachers' beliefs about using web 2.0 technologies in k-12 classroom. *Computers & Education*, 59(3):937–945, 2012.

[11] H. Wang, X. He, M.-W. Chang, Y. Song, R. W. White, and W. Chu. Personalized ranking model adaptation for web search. In *ACM SIGIR*, pages 323–332, 2013.