

YouUnderstood.me? Readability-based retrieval of online reading 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. Students can often get discouraged because the contents they retrieve are outside their comprehension level. For educators, finding materials for curriculum development that suit their students' reading abilities can also be challenging. In this paper, we present *YouUnderstood.me*, a web application that makes use of natural language processing, machine learning and information retrieval techniques to help both students and educators in the process of finding materials that fit the reading skills of each individual student in a faster and more efficient way. *YouUnderstood.me* combines: (1) a search interface that by putting together a search engine and a readability formula, permits the fast retrieval of documents from different sources, (2) a tracking system that makes possible to track the reading level of individual students and (3) an analysis tool that enables educators to analyze the reading level of specific reading materials found outside *YouUnderstood.me*.

CCS Concepts

•Information systems → Personalization; Search interfaces; •Applied computing → Interactive learning environments; Learning management systems;

Keywords

Search interfaces; Filtering; Readability; Educational tools

K-12 students use the internet on a daily basis to seek materials that can help them with their academic tasks, such as finding information for a class presentation or discovering the meaning of a new word. For this purpose, they often turn to search engines and online catalogs to retrieve reading materials that can satisfy their information needs, including news articles, books or term definitions. However, students can often get discouraged because the contents they retrieve

are outside their comprehension level, whether being too easy or too difficult for them to understand. Therefore, providing them ways to seek for adequate materials they can actually comprehend is imperative.

In the academic environment students are not the only ones facing the problem of locating adequate reading materials that simultaneously match their information needs and reading abilities. Educators also face several challenges when looking for materials for their classes, which makes them spend a significant amount of time doing so. 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. This entails a personalization need that the instructor has to handle on a daily basis. However with the high number of students in 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 reading materials. The system is centered on the use of readability formulas that together with a search engine makes looking for levelled reading material fast and efficient. *YouUnderstood.me* lets students log in in the application, which keeps track of feedback they give on retrieved materials (too easy/OK/too complex). This 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 score from: (i) commercial search engines, such as Google, (ii) public data sources, such as Wikipedia and (iii) local resources, such as the catalogues of a school library.

Besides the search interface, the instructors also have access to an analysis page, where they can submit texts they found outside the application for determining 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.

The novelty of *YouUnderstood.me* lays in how its different submodules are combined in order to create an application that becomes helpful in an academic environment. 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 cooper-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WOODSTOCK '97 El Paso, Texas USA

© 2016 ACM. ISBN 123-4567-24-567/08/06...\$15.00

DOI: 10.475/123_4

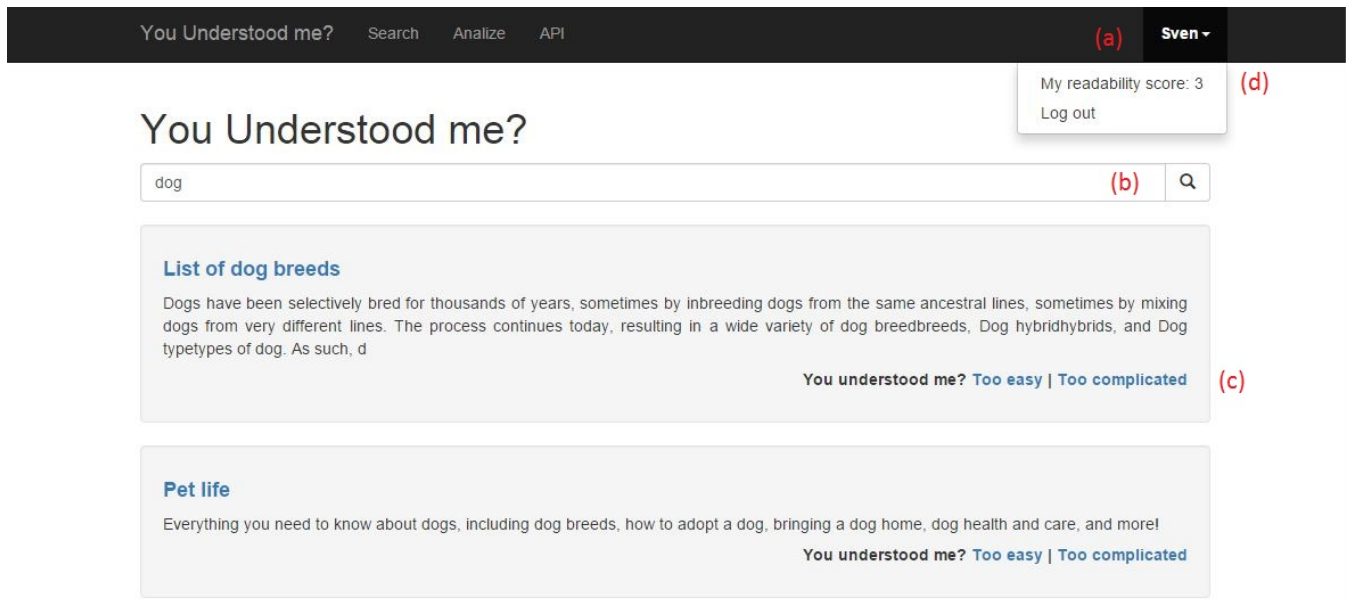


Figure 1: Screenshot of the search interface for students

ation to speed-up the process of reading material selection.

1. 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 features, such as, the average number of words per sentence or the average length of terms, to more complex ones [5][1] [4], 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. Therefore, different works have been done in more specialized contexts such as book [2][9] or web page retrieval [Ref], 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 assess text complexity are the following:

- **External metrics.** *YouUnderstood.me* is compatible with the most popular readability metrics among the American education centers and libraries, such as AR¹ or Lexile². The aforementioned metrics are widely used for measuring 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 supported by *YouUnderstood.me*.
- **MRAS.** MRAS [7] (Multilingual Readability assessment system) is a state of the art readability assessment system, that is capable of detecting the input language of a text on the fly and providing a readability score for it. 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 fig. 1) which provides several features that are aimed at helping them in the process of seeking for 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 inserts is first treated with an ad-hoc search intent module that is capable of treating issues that usually arise while processing children queries and that usually lead children too retrieve poor results, such as misspelling, children popular culture terms or too long queries. Processed queries are submitted to a search engine, which is different depending on the source required. Currently two methods of search are implemented: one that makes use of the Google Search API and another one that based on the Apache Lucene³ framework.

Documents retrieved by the search engine are analyzed and filtered based on the readability requirements of each

¹www.acceleratelearning.com

²www.lexile.com

³lucene.apache.org/

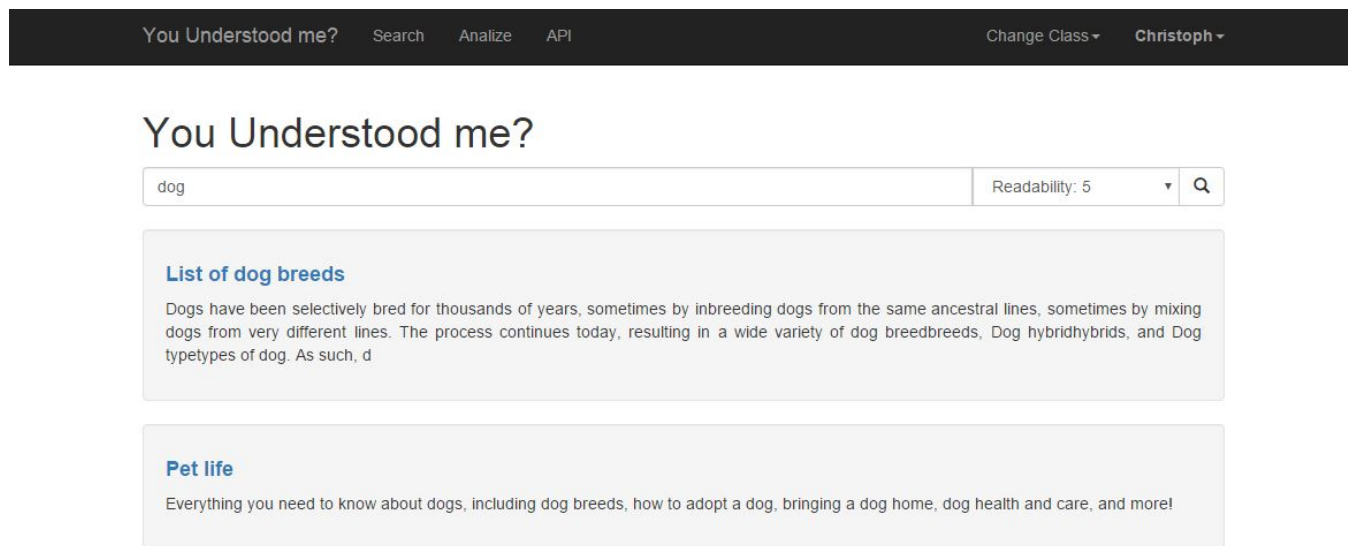


Figure 2: Screenshot of the search interface for educators

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 that the readability prediction needs to be fast, forcing the application to use a lightweight readability formula such as Flesch or FOG. However, when using Lucene, the 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 give feed-back on it (c), helping the system improve the filtering for future searches. This feed back is given by means of a three option question which permits the student state whether the reading material was too easy, OK, or too complicated. This information is provided to the tracking system to create predictions about individual users' readability skills.

The **tracking system** 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 keep track of the progress he is making.

2. INTRODUCTION

Educators can also make use of the **search interface**, which is adapted to their use (fig. 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 intend module, given that the 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 feed-back

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 (fig. 4) that allows them to analyze snippets of text self-outlined. This tool allow to choose the readability formula used for the analysis, among most of the formulas that have already been mentioned.

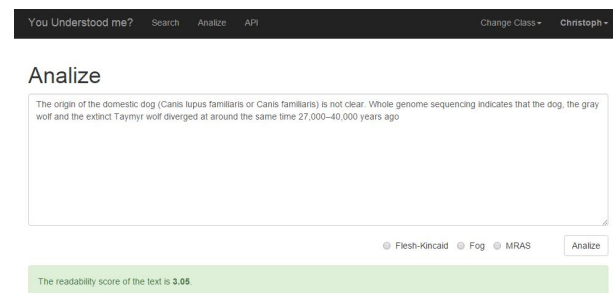


Figure 4: Screenshot of analysis page

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 the 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

TODO

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, as well as ..., making it extensible and versatile. It provides an environment tailored to different types of users: students




You Understood me? Search Analyze API Change Class Christoph						
Student list						
Id	Last Name	Name	Readability	Last connection	Image	Actions
22	Lorrin	Peter	5	January 24, 2016 at 9:00 PM		View Edit
22	Anderson	Annie	3	January 23, 2016 at 11:00 PM		View Edit
22	Connor	Mike	5	January 24, 2016 at 9:00 PM		View Edit

Figure 3: Screenshot of student list for educator

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 XXX, i.e., a search interface, which aids educators in locating materials tailored to individual students, a readability analysis tool, which can help them in the 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 the 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.

TODO FUTURE WORK

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 *Proceedings of the second workshop on speech and language processing for assistive technologies*, pages 73–83. Association for Computational Linguistics, 2011.
- [2] J. Denning, M. S. Pera, and Y.-K. Ng. A readability level prediction tool for k-12 books. *Journal of the Association for Information Science and Technology*, 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 *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 466–477. Association for Computational Linguistics, 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 *Proceedings of COLING*, volume 2014, 2014.
- [6] R. Gunning. {The Technique of Clear Writing}. 1952.
- [7] I. Madrazo and S. Pera. Steps towards multilingual readability assessment. In press.
- [8] J. V. Oakhill and K. Cain. The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific studies of reading*, 16(2):91–121, 2012.
- [9] M. S. Pera and Y.-K. Ng. Automating readers' advisory to make book recommendations for k-12 readers. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 9–16. ACM, 2014.
- [10] R. D. Robinson, M. C. McKenna, and J. M. Wedman. Issues and trends in literacy education. 2000.