YouUnderstood.me? Readability based retrieval of reading materials for students and educators

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

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

ABSTRACT

K-12 students make use of online resources to fulfill their academic information needs on a daily basis. However, they can often get discouraged because the contents they retrieve are outside their comprehension level, whether being too easy or too difficult for them to read. At the same time, educators also find several challenges locating materials for curriculum development that suit the students' reading skills. Those reasons, make both students and educators spend a reasonably large amount of time seeking for adequate materials.

In this paper, we present a web application that makes use of techniques, coming from natural language processing, machine learning and information retrieval areas, 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. For this purpose, the web application we present, combines: (1) a search interface that by combining a search engine and a readability formula, permits the fast retrieval of documents from different sources, (2) a readability tracking system that enables both types of users to see how the reading skills of the student are evolving with time and (3) an analysis tool that enables educators to analyse materials from outside the application for determining their complexity level.

CCS Concepts

•**Information systems** → **Personalization; Search interfaces;** *Clustering and classification;* •**Computing methodologies** → **Natural language processing;**

Keywords

Search engines; Filtering; Readability assessment; Student tracking

WOODSTOCK '97 El Paso, Texas USA

DOI: 10.475/123_4

1. INTRODUCTION

K-12 students make use of the internet in a daily basis to fulfill their information needs for completing academic tasks, such as finding information for a class presentation or discovering the meaning of a new word. For this purpose, they often turn search engines and online available catalogues for retrieving reading materials that can satisfy their information needs, including news articles, books or term definitions. However, they can often get discouraged because the contents they retrieve are outside their comprehension level, whether being too easy or too difficult for them to read. And why is that a problem?

In the academic environment students are not the only ones facing the problem of locating adequate reading materials that simultaneously match the information needs and reading abilities of an individual. For example, even in a same grade class, students' reading skills can differ significantly, so not all students in the same class can be provided with same texts. This supposes 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. Due to the mentioned fact, instructors spend a significant amount of their time seeking adequate materials for their students, a time that could be reduced with the help of an specialized application.

YouUnderstood.me is a web application oriented to both instructors and students , that aims at helping them in the process of finding reading materials. The system is centred 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 web application, which keeps track of the materials read by individual students and their feed-back for each or the read materials(too easy/OK/too complex). This enables the application to make predictions about the readability score for each student. A score that, both the student and the teacher are able too see in order to keep track of how it evolves. Furthermore, the application contains a search engine that allows both students and educators to seek materials filtered by a their readability score from: (i) commercial search engines, such as google, (ii) public data sources, such as Wikipedia, AR[1] or Lexile[2] and (iii) local resources, such as the catalogues of a school library.

[1]http://www.acceleratelearning.com
[2]http://lexile.com

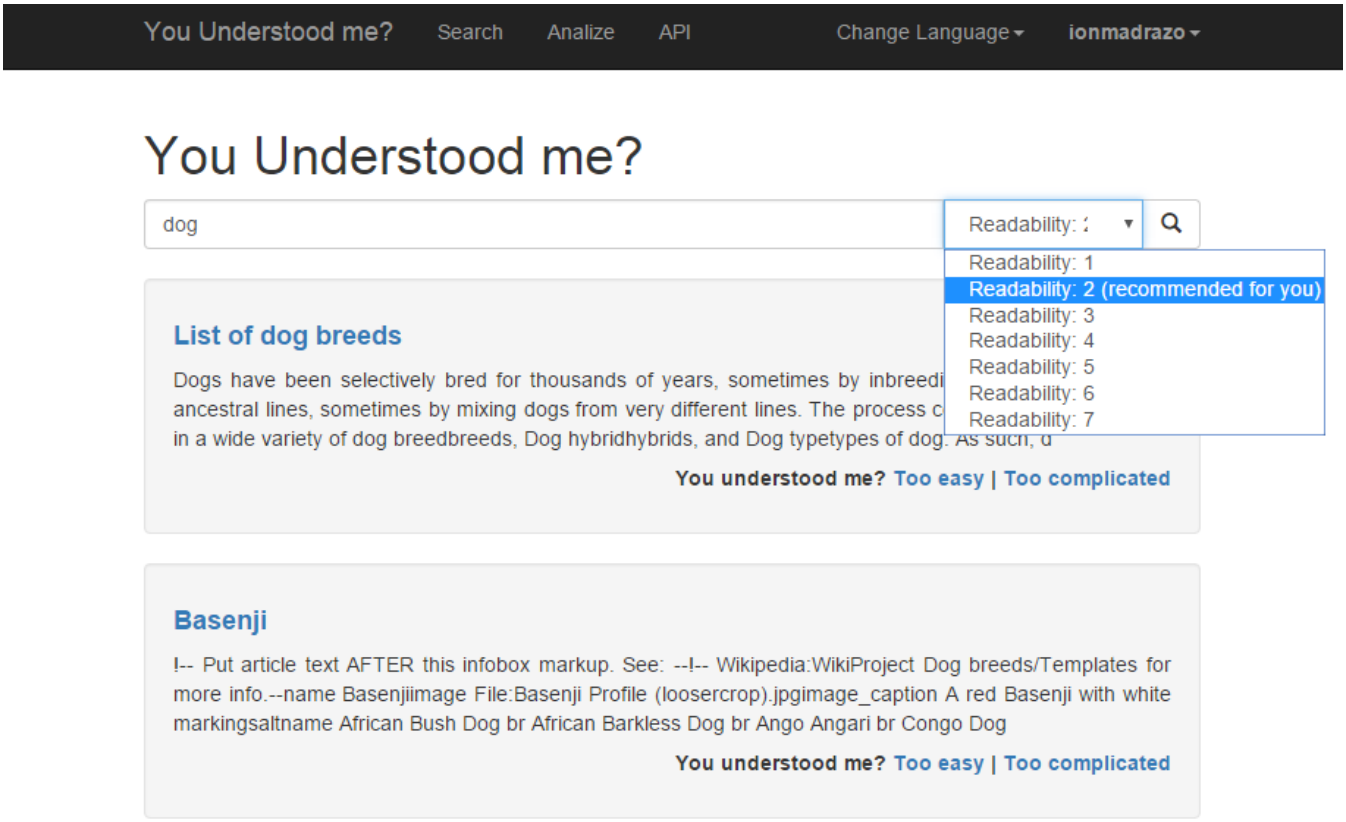

Figure 1: Screenshot of the search interface

Finally, the instructors also have access to an analysis page, where they can submit texts they found outside the application for determining their readability score, being able to choose from a wide range of readability formulas provided within the application. This tool, together with the track of readability scores of each students, helps teachers make sure the reading materials they found are adequate or not for the class.

2. READABILITY ASSESSMENT

Whether when a student is searching information for completing an assignment or when an educator uses the application for finding material for his course, a text's complexity needs to be determined.

Different approaches have been followed in the literature for determining a text's complexity or readability. Most approaches, orient their work to determining the readability of a general text. Those systems vary, from the very simple ones [Flesch, 1948], 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 [**?**][Dell'Orletta et al., 2011] [François and Fairon, 2012], which are mostly based, on supervised learning techniques and features extracted using Natural language processing. However, those tools have shown to be of small use in contexts where the text of the reading material has reduced accessibility, both because the texts is not publicly accessible or because it shows a structure that is not as simple to tackle . Therefore, different works have been done in more specialized contexts such as book[Denning et al., 2015][Pera and Ng, 2014] or web page retrieval[Ref], where the systems presented made use other features apart from the ones in the text.

YouUnderstood.me can make use of different readability assessment metrics and resources at the same time, aiming to be able to handle a more diverse amount of reading materials. The methods which with *YouUnderstood.me* can assess a reading material readability are the following:

- **External metrics.** *YouUnderstood.me* is compatible with the most popular readability metrics among the American education centers and libraries, such as the ones that are used in material catalogs like AR or Lexile. The compatibility with those types of metrics allows the application to be able to retrieve books, which have been historically an issue for readability formulas, because of the inability to get access to the contents of copyrighted material [Denning et al., 2015].
- **Traditional formulas.** Historically used by teachers for manually determining the readability level of a reading material, traditional formulas such as Flesh [Flesch, 1948], Fog[Gunning, 1952] and Flesh-Kincaid [Flesch, 1948], are supported by *YouUnderstood.me*. These formulas enable educators compare new materials with the old, hand-way measured materials they may have stored during their teaching career.
- **MRAS.** MRAS[Madrazo and Pera, ress] (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, at the moment, makes use of more than a hundred features for learning and prediction. The features are extracted using Natural Language Processing tools, such as a tokenizer, a part of speech tagger and different semantic analyzers. Those features, are used to train a model that will later be used for predicting the readability score for new materials.

3. SEARCH INTERFACE

4. TRACKING STUDENTS

5. USER EXPERIENCE

- **Analyzing materials.** An analysis page is provided to instructors, so that they can make use of the readability assessment algorithm [Madrazo and Pera, ress] with materials from outside the application. This page provides an input form where the instructor can submit the material, and receive a readability level prediction for the it.

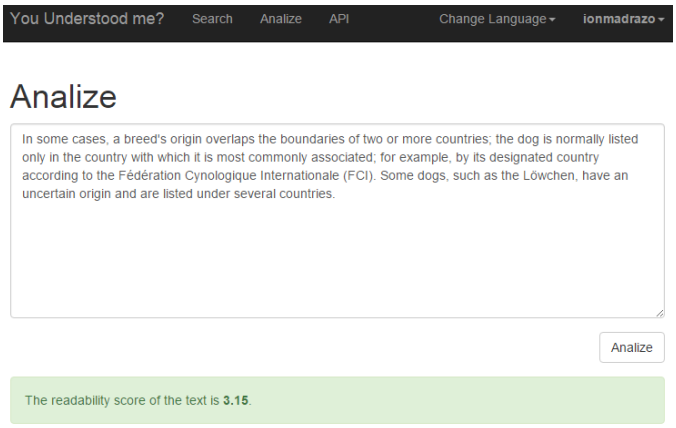

Figure 2: Screen shot of analysis page

- **Looking for new materials.** A material search page is provided so that the user can insert text queries for looking for materials. The user can select a certain readability level he wants to look for, of leave it blank letting the application choose the best level for the logger user. Furthermore, the material source can be selected, depending on the user needs. In case of students, they can select a material for reading and provide feed-back on it. This feed-back will be used for tracking purposes.

- **Tracking students.** The educator can see a table with data about all his students, where the number of materials read and the readability score for each student are provided. The instructor can go deeper if needed, and see each of the materials each students has read and the individual readability scores for each material, as well as, the feed-back given by the student. The data is also presented in a summarized way, so that the educator can see the average, maximum and minimum reading skills of the students in class.

6. FUTURE WORK

7. MISSING IDEAS

- The application is fully multilingual.

8. REFERENCES

[Dell'Orletta et al., 2011] Dell'Orletta, F., Montemagni, S., and Venturi, G. (2011). 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.

[Denning et al., 2015] Denning, J., Pera, M. S., and Ng, Y.-K. (2015). A readability level prediction tool for k-12 books. *Journal of the Association for Information Science and Technology*.

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

[François and Fairon, 2012] François, T. and Fairon, C. (2012). 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.

[Gonzalez-Dios et al., 2014] Gonzalez-Dios, I., Aranzabe, M. J., de Ilarraza, A. D., and Salaberri, H. (2014). Simple or complex? assessing the readability of basque texts. In *Proceedings of COLING*, volume 2014.

[Gunning, 1952] Gunning, R. (1952). {The Technique of Clear Writing}.

[Madrazo and Pera, ress] Madrazo, I. and Pera, S. (In press). Steps towards multilingual readability assessment tool.

[Pera and Ng, 2014] Pera, M. S. and Ng, Y.-K. (2014). 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.

[Robinson et al., 2000] Robinson, R. D., McKenna, M. C., and Wedman, J. M. (2000). Issues and trends in literacy education.