

# 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 and educators make use of online resources to fulfill their academic information needs on a daily basis. Unfortunately, they are forced to spend a large amount of time seeking for adequate materials. In the case of the students, 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. For educators, finding materials for curriculum development that suit the 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 joining a search engine and a readability formula, permits the fast retrieval of documents from different sources, (2) a readability tracking system that enables the aforementioned stakeholders to see which the reading skills of individual students are and (3) an analysis tool that enables educators to analyze specific reading materials found outside *YouUnderstood.me*.

## CCS Concepts

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

## Keywords

Search engines; Filtering; Readability assessment; Student tracking

## 1. INTRODUCTION

K-12 students make use of the internet in a daily basis seeking 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 available catalogs 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, thus, helping them 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 the information needs and reading abilities of an individual. Educators also face several challenges when looking for materials for their classes, making them spend a significant amount of their time doing so. 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.

*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 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 feed-back they have given for each of the read materials (too easy/OK/too complex). This enables the application to make predictions about the readability score for each student, and give both the student and his educator, the possibility to see an updated readability score that can be helpful for speeding-up the process of searching adequate materials. Furthermore, the application contains a search engine that allows both students and educators 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.

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

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

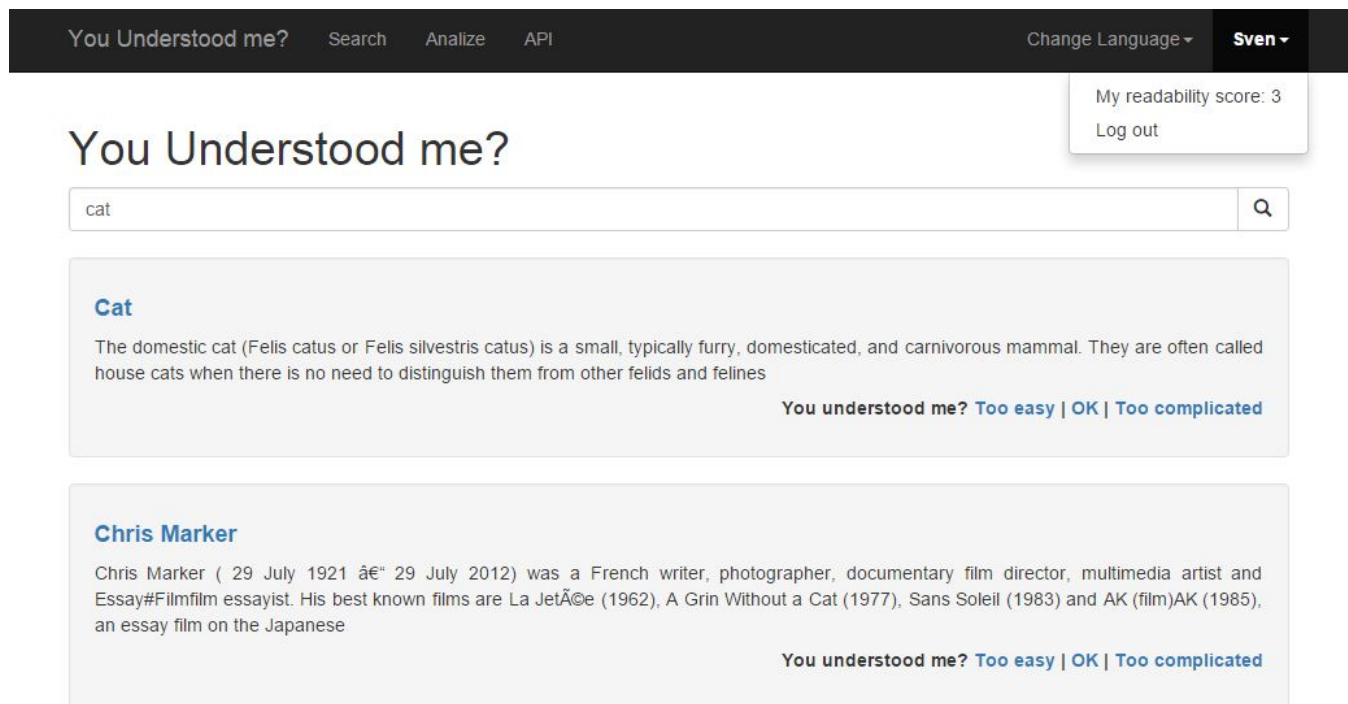


Figure 1: Screenshot of the search interface for students

sure the reading materials they found are adequate or not for the class.

The novelty of *YouUnderstood.me* lays in how the different modules 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 cooperation in order to speed-up the process of reading material selection.

## 2. YOUUNDERSTOOD.ME

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 in order to evaluate if the material fits the needs of readability level.

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 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 [Gonzalez-Dios et al., 2014][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 with which *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, 2015] (Multilingual Readability assessment system) is a state of the

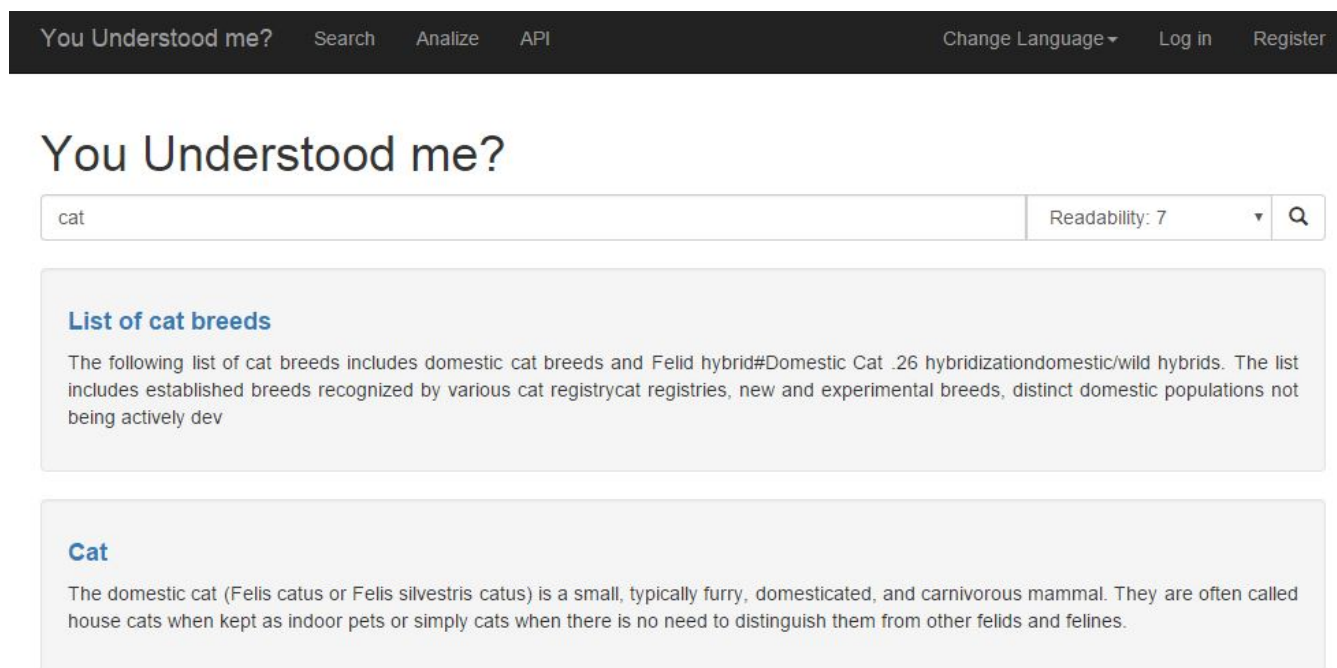


Figure 2: Screenshot of the search interface for educators

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.

The stakeholders of *YouUnderstood.me* can benefit of the aforementioned readability scores in different ways. On the one side, **students** can make use of a **search engine** (fig. 1) which provides several features that are aimed at helping the student in the process of seeking an adequate reading material. In order to use it, the student needs to be logged in (a), which permits the application personalize the experience the student is having by different means.

Each query a student inserts is first treated with **OSIM** (**Our search intent module**)[Ref], an ad-hoc developed search intent module for children. OSIM is capable of treating issues that usually arise while processing children queries, such as misspelling, children popular culture terms or too long queries [Ref]. Issues that usually lead children to get poor results while using a common search engine.

Once the query is treated and converted into a new query, it is submitted to the corresponding search tool. Currently two methods of search are implemented: a method that makes use of the Google Search API and a method based the lucene framework.

Once results are retrieved by the corresponding search engine, they are analyzed for their readability and filtered by 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, the documents need to be analyzed on the fly when using Google, since the documents are not in possession of the application. This supposes 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 recommendations 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 in order 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 conscious 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), permitting each student keep track of the progress he is making.

**Educators** can also make use of the **search engine** interface, which is adapted to their use (fig. 2). The educator

# Student list







Id	Last Name	Name	Readability	Last connection	Image	Actions
22	Lorrin	Peter	5	January 24, 2016 at 9:00 PM		<a href="#">View</a> <a href="#">Edit</a>
15	Into	Vahit	3	January 23, 2016 at 11:00 PM		<a href="#">View</a> <a href="#">Edit</a>
17	Connor	Mike	5	January 24, 2016 at 9:00 PM		<a href="#">View</a> <a href="#">Edit</a>
19	Anderson	Billie	2	January 23, 2016 at 11:00 PM		<a href="#">View</a> <a href="#">Edit</a>
5	Halderson	Jason	7	January 24, 2016 at 9:00 PM		<a href="#">View</a> <a href="#">Edit</a>
21	Richarson	Marissa	4	January 23, 2016 at 11:00 PM		<a href="#">View</a> <a href="#">Edit</a>

Figure 3: Screenshot of student list for educator

view of the search engine does not require the educator to be logged in, giving the user himself the possibility of personalizing the searches. The search engine no longer makes use of the search intend module, given that the educators 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 educator 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 make use of the material analysis page (fig. 4). This page, permits educators analyze reading materials they cannot find using the provided search engine, i.e. when materials are self created. This tool allow to choose the readability formula used for the analysis, among all the formulas that have already been mentioned.

You Understood me?
 Search
 Analyze
 API
 Change Language
 ionmadrazo

### Analyze

In some cases, a breed's origin overlaps the boundaries of two or more countries; the dog is normally listed only in the country with which it is most commonly associated; for example, by its designated country according to the Fédération Cynologique Internationale (FCI). Some dogs, such as the Löwchen, have an uncertain origin and are listed under several countries.

Analyze

The readability score of the text is 3.15.

Figure 4: Screen shot of analysis page

Finally, a student tracking tool is provided for educators (fig. 3). This tool provides a fast way to know the indi-

vidual 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 feed-back that the student gave for each of the read materials.

### 3. CONCLUSION AND FUTURE WORK

### 4. 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.