

Real-Time Web Text Classification and Analysis of Reading Difficulty

Eleni Miltsakaki

Graduate School of Education
University of Pennsylvania,
Philadelphia, PA 19104, USA.
elenimi@seas.upenn.edu

Audrey Troutt

Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
atroutt@seas.upenn.edu

Abstract

The automatic analysis and categorization of web text has witnessed a booming interest due to the increased text availability of different formats, content, genre and authorship. We present a new tool that searches the web and performs in real-time a) html-free text extraction, b) classification for thematic content and c) evaluation of expected reading difficulty. This tool will be useful to adolescent and adult low-level reading students who face, among other challenges, a troubling lack of reading material for their age, interests and reading level.

1 Introduction

According to the National Center for Education Statistics, 29% of high school seniors in public schools across America were below basic achievement in reading in 2005 (U.S. Department of Education 2005). Once these students enter high school, their reading problems, which began much earlier in their education, are compounded by many factors including a lack of suitable reading material for their age, interests and reading level. Most material written at a lower reading level is designed for much younger students; high-school students find it boring or embarrassing. On the other hand material designed for older students, while probably more interesting, is incomprehensible to such a student and leads to frustration and self-doubt. The internet is a vast resource for potential reading material and is often utilized by educators in the classroom, but it is not currently possible to filter the results of a search

engine query by levels of readability. Instead, the software that some schools have adopted restricts students to lists and directories of hand-selected educational sites. This severely limits the content available to students and requires near-constant maintenance to keep current with new information available on the web.

We are developing a new system, Read-X, that searches the web and performs in real-time a) html-free text extraction, b) classification for thematic content and c) evaluation of expected reading difficulty. For the thematic classification task we collected a manually labeled corpus to train and compare three text classifiers. Our system is part of larger research effort to improve existing readability metrics by taking into account the profile of the reader. As a first step in this direction, we computed vocabulary frequencies per thematic area. We use these frequencies to predict unknown words for the reader relative to her familiarity with thematic areas (Toreador). These tools (Read-X and Toreador) will be useful to adolescent and adult low-level reading students who face, among other challenges, a troubling lack of reading material for their age, interests and reading level.

The remainder of the paper is organized as follows: first we will describe our motivation for creating Read-X and Toreador, which is based on studies that show that older struggling readers can make improvements in literacy and that those improvements can have a profound impact on their lives. Next we will describe existing technologies for literacy improvement and research related to our current project. Finally, we will give a detailed description

of Read-X and Toreador, including our methods of evaluating the readability of texts, thematically classifying the texts and modeling reader profiles into readability predictions, before concluding with an outline of future work.

2 Educational motivation

Low reading proficiency is a widespread problem evident in the performance of adolescents in U.S. schools. The National Center for Education Statistics (NCES) in 2005, the latest year for which data is available, reports that only 29% of eight graders in the United States achieved proficient or above reading, meaning the remaining 71% of students had only part of the reading skills needed for proficient work at their level or less (Snyder et al., 2006). (Hasselbring and Goin, 2004) reported that "as many as 20 percent of 17-year-olds have been estimated to be functionally illiterate, and 44 percent of all high-school students have been described as semi-literate". Reading below grade level is a serious problem for adolescents as it may hinder comprehension of textbooks and classroom materials in all fields. (Denti, 2004) mentions that "most high school textbooks are written at the tenth through twelfth grade levels with some textbooks used for U. S. government written at the seventeenth grade level". Reading skills are tied to academics success and are highly correlated with with "higher income and less unemployment, increased access to lifelong learning, greater amounts of personal reading for pleasure, and increased civic participation" (Strucker et al., 2007).

Recent research has shown that it is possible to identify adult literacy students on the brink of achieving reading fluency in order to provide them with concentrated instruction, dramatically improving their chances of attaining a high quality of life (Strucker et al., 2007). (Weinstein and Walberg, 1993) studied the factors related to achievement in reading and found that "frequent and extensive engagement in literacy-promoting activities as a young adult was associated with higher scores on literacy outcomes (independent of earlier-fixed characteristics and experiences)," which implies that through ample reading exercise students can achieve literacy regardless of their background.

The current and future versions of the system that we are developing uses natural language processing techniques to provide learning tools for struggling readers. The web is the single most varied resource of content and style, ranging from academic papers to personal blogs, and is thus likely to contain interesting reading material for every user and reading ability. The system presented here is the first to our knowledge which performs in real time a)keyword search, b)thematic classification and c)analysis of reading difficulty. We also present a second system which analyzes vocabulary difficulty according to reader's prior familiarity with thematic content.

3 Related work

In this section we discuss two main systems that are most closely related to our work on text classification and analysis of readability.

NetTrekker is a commercially available search tool especially designed for K-12 students and educators.¹ NetTrekker's search engine has access to a database of web links which have been manually selected and organized by education professionals. The links are organized thematically per grade level and their readability level is evaluated on a scale of 1-5. Level 1 corresponds to reading ability of grades 1-3 and 5 to reading ability of grades 11-13. NetTrekker has been adopted by many school districts in the U.S., because it offers a safe way for K-12 students to access only web content that is age appropriate and academically relevant. On the other hand, because the process of web search and classification is not automated, it is practically impossible for NetTrekker to dynamically update its database so that new material posted on the web can be included. However, NetTrekker's manual classification of web links is a valuable resource of manually labeled data. In our project, we use this resource to build labeled dataset for training statistical classifiers. We discuss the construction and use of this corpus in more detail in Section 5.1).

The REAP tutor, developed at the Language Technologies Institute at Carnegie Mellon, is designed to assist second language learners to build new vocabulary and facilitates student specific practice sessions (Collins-Thompson and Callan, 2004), (Heilman et

¹ Available at <http://www.nettrekker.com>.

al., 2006). The tutor allows the user to search for textual passages as well as other text retrieved from the web that contains specific vocabulary items. The educational gain for students practicing with the tutor has been shown in several studies (e.g., (Heilman et al., 2006)). Like NetTrekker, REAP retrieves and classifies web text off-line. Unlike, Nettrekker, however, textual analysis is automated. REAP's information retrieval system (Collins-Thompson and Callan, 2004) contains material from about 5 million pages gathered with web crawling methods. The data have been annotated and indexed off-line. Annotations include readability level computed with an earlier version of the method developed by (Heilman et al., 2007), (Heilman et al., 2006) described below, rough topic categorizations (e.g., fiction, non-fiction) and some elements of grammatical structure (e.g., part-of-speech tagging).

(Heilman et al., 2007) experiment with a system for evaluation of reading difficulty which employs both grammatical features and vocabulary. The grammatical features built in the model were identified from grammar books used in three ESL levels. (Heilman et al., 2007) find that while the vocabulary model alone outperformed the grammar-based model, the combined model performed best. All models performed better in English text and less well in ESL text. It would be very interesting to integrate this system with Read-X and evaluate its performance.

To address issues specific to struggling readers, (Hasselbring and Goin, 2004) developed the Peabody Literacy Lab (PLL), a completely computer-based program, using a variety of technologies to help students improve their ability to read. We will not elaborate further on this work because the PPL's focus is not in developing new technologies. PLL develops experimental programs using existing technologies.

4 Read-X project overview

In the Read-X project, we have developed two tools which are currently independent of each other. The first tool Read-X, performs a web search and classifies text as detailed in (5.1). The second tool Toreador, analyzes input text and predicts vocabulary difficulty based on grade or theme-specific vocabulary

frequencies. The vocabulary predicted to be unfamiliar can be clicked on. This action activates a dictionary look-up search on Wordnet whose display is part of the tool's interface. More details and screenshots are given in (??).

5 Description of Read-X

Below we describe in detail the technical components of Read-X: internet search, text extraction and analysis of readability.

5.1 Read-X: Web search and text classification

Internet search. Read-X performs a search of the internet using the Yahoo! Web Services. When the search button is clicked or the enter key depressed after typing in a keyword, Read-X sends a search request to Yahoo! including the keywords and the number of results to return and receives results including titles and URLs of matching websites in an XML document. The Yahoo! Web Service is freely available for non-commercial use with a limit of 5000 requests per day. If Read-X is deployed for use by a wide number of users, it may be necessary to purchase the ability to process more requests with Yahoo or another search engine. Read-X is currently available at <http://net-read.blogspot.com>.

Text extraction. Read-X then retrieves the html, xml, doc or PDF document stored at each URL and extracts the human-readable text.² text is extracted from html and xml documents using the scraper provided by Generation Java by Henri Yandell, see www.generationjava.com. The Microsoft Word document scraper is part of the Apache Jakarta project by the Apache Software Foundation, see www.apache.org. The PDF scraper is part of the Apache Lucene project, see www.pdfbox.org. All three of these external tools are available under a common public license as open source software under the condition that any software that makes use of the tools must also make the source code available to users.

²Being able to identify appropriate web pages whose content is reading material and not "junk" is a non-trivial task. (Petersen and Ostendorf, 2006) use a classifier for this task with moderate success. We "read" the structure of the html text to decide if the content is appropriate and when in doubt, we err on the side of throwing out potentially useful content.

Readability analysis. For printed materials, there are a number of readability formulas used to measure the difficulty of a given text; the New Dale-Chall Readability Formula, The Fry Readability Formula, the Gunning-Fog Index, the Automated Readability Index, and the Flesch Kincaid Reading Ease Formula are a few examples. Usually these formulas count the number of syllables, long sentences, or difficult words in randomly selected passages of the text. To automate the process of readability analysis, we chose three Readability algorithms: Lix, Rix, see (Anderson, 1983), and Coleman-Liau, (Coleman and Liau, 1975), which were best suited for fast calculation and provide the user with either an approximate grade level for the text or a readability classification of very easy, easy, standard, difficult or very difficult. When each text is analyzed by Read-X the following statistics are computed: total number of sentences, total number of words, total number of long words (seven or more characters, and total number of letters in the text. Below we describe how each of the three readability scores are calculated using these statistics. Steps taken to develop more sophisticated measures for future implementations are presented in Section 7).

Lix readability formula: The Lix readability algorithm distinguishes between five levels of readability: very easy, easy, standard, difficult, or very difficult. If W is the number of words, LW is the number of long words (7 or more characters), and S is the number of sentences, then the Lix index is $LIX = W/S + (100 * LW) / W$. An index of 0-24 corresponds to a very easy text, 25-34 is easy, 35-44 standard, 45-54 difficult, and 55 or more is considered very difficult.

Rix readability formula: The Rix readability formula consists of the ratio of long words to sentences, where long words are defined as 7 or more characters. The ratio is translated into a grade level as indicated in Table (1).

Coleman-Liau readability formula: The Coleman-Liau readability formula is similar to the Rix formula in that it gives the approximate grade level of the text. Unlike the Lix and Rix formulas, the Coleman-Liau formula requires the random selection of a 100 word excerpt from the text. Before the grade level can be calculated, the cloze percent must be estimated for this selection. The

Ratio	GradeLevel
7.2 and above	College
6.2 and above	12
5.3 and above	11
4.5 and above	10
3.7 and above	9
3.0 and above	8
2.4 and above	7
1.8 and above	6
1.3 and above	5
0.8 and above	4
0.5 and above	3
0.2 and above	2
Below 0.2	1

Table 1: Rix translation to grade level

Classifier	Supercategories	Subcategories
Naive Bayes	66%	30%
MaxEnt	78%	66%
MIRA	76%	58%

Table 2: Performance of text classifiers.

cloze percent is the percent of words that, if deleted from the text, can be correctly filled in by a college undergraduate. If L is the number of letters in the 100 word sample and S is the number of sentences, then the estimated cloze percent is $C = 141.8491 - 0.214590 * L + 1.079812 * S$. The grade level can be calculated using the Coleman-Liau formula, where grade level is $-27.4004 * C + 23.06395$. In the SYS display we round the final result to the nearest whole grade level.

6 Text classification

The automated classification of text into predefined categories has witnessed strong interest in the past ten years. The most dominant approach to this problem is based on machine learning techniques. Classifiers are built which learn from a pre-labeled set of data the characteristics of the categories. The performance of commonly used classifiers varies depending on the data and the nature of the task. For the text classification task in Read-X, we a) built a corpus of pre-labeled thematic categories and b) compared the performance of three classifiers to evaluate their per-

formance on this task.

We collected a corpus of approximately 3.4 million words and organized it into two sets of labeling categories. We hand collected a subset of labels (most appropriate for a text classification task) from the set of labels used for the organization of web text in NetTrekker (see 3). We retrieved text for each category by following the listed web links in NetTrekker and manually extracting text from the sites. Our corpus is organized into a small hierarchy, with two sets of labels: a) labels for supercategories and b) labels for subcategories. There are 8 supercategories (Arts, Career and business, Literature, Philosophy and religion, Science, Social studies, Sports and health, Technology) and 41 subcategories (e.g., the subcategories for Literature are Art Criticism, Art History, Dance, Music, Theater). Subcategories are a proper subset of supercategories but in the classification experiments reported below the classifiers trained independently in the two data sets.

We trained three classifiers for this task: a Naive Bayes classifier, a Maximum Entropy classifier and MIRA, a new online learning algorithm that incorporates a measure of confidence in the algorithm (for details see (Crammer et al., 2008)).³ The performance of the classifiers trained on the supercategories and subcategories data is shown in Table (2). All classifiers perform reasonably well in the supercategories classification task but are outperformed by the MaxEnt classifier in both the supercategories and subcategories classifications. The Naive Bayes classifier performs worst in both tasks. As expected, the performance of the classifiers deteriorates substantially for the subcategories task. This is expected due to the large number of labels and the small size of data available for each subcategory. We expect that as we collect more data the performance of the classifiers for this task will improve. In an earlier implementation of Read-X, thematic classification was a coarser three-way classification task (literature, science, sports). In that implementation the MaxEnt classifier performed at 93% and the Naive Bayes classifier performed at 88% correct. In future implementations of the tool, we will make available

³We gratefully acknowledge MALLET, a collection of statistical NLP tools written in Java, publicly available at <http://mallet.cs.umass.edu> and Mark Dredze for his help installing and running MIRA on our data.

all three levels thematic classification.

6.1 Runtime and interface

The first implementation of Read-X, coded in Java, has been made publicly available. The jar file is called from the web through a link and runs on Windows XP or Vista with Java Runtime Environment 6 and internet connection. Search results and analysis are returned within a few seconds to a maximum of a minute or two depending on the speed of the connection. The Read-X interface allows the user to constrain the search by selecting number of returned results and level of reading difficulty. A screenshot of Read-X (cropped for anonymity) is shown in Figure (1). The rightmost column is clickable and shows the retrieved html-free text in an editor. From this editor the text can be saved and further edited on the user's computer.

7 Description of Toreador

The analysis of reading difficulty based on standard readability formulas gives a quick and easy way to measure reading difficulty but it is problematic in several ways. First, readability formulas compute superficial features of word and sentence length. It is easy to show that such features fail to distinguish between sentences which have similar word and sentence lengths but differ in ease of interpretation. Garden path sentences, bountiful in the linguistic literature, demonstrate this point. Example (1) is harder to read than example (2) although the latter is a longer sentence.

- (1) She told me a little white lie will come back to haunt me.
- (2) She told me that a little white lie will come back to haunt me.

Secondly, it is well known that there are aspects of textual coherence such as topic continuity and rhetorical structure which are not captured in counts of words and sentences (e.g., (Higgins et al., 2004), (Miltsakaki and Kukich, 2004))

Thirdly, readability formulas do not take into account the profile of the reader. For example, a reader who has read a lot of literary texts will have less difficulty reading new literary text than a reader, with a similar educational background, who has never read

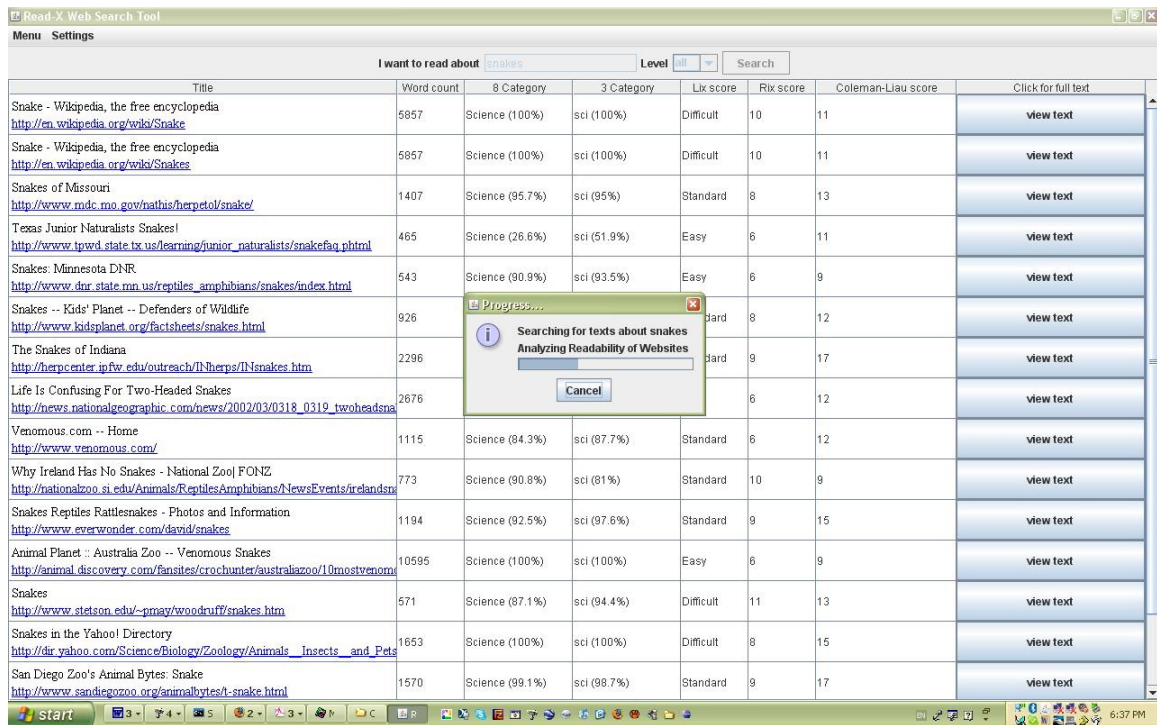


Figure 1: Search results and analysis of readability

any literature. In this section, we discuss the first step we have taken towards making more reliable predictions on text readability given the profile of the reader.

Readers who are familiar with specific thematic areas, are more likely to know vocabulary that is recurring in these areas. So, if we have vocabulary frequency counts per thematic area, we are in a better position to predict difficult words for specific readers given their reading profiles. Vocabulary frequency lists are often used by test developers as an indicator of text difficulty, based on the assumption that less frequent words are more likely to be unknown. However, these lists are built from a variety of themes and cannot be customized for the reader. We have computed vocabulary frequencies for all supercategories in the thematically labeled corpus. The top 10 most frequent words per supercategory are shown in Table (3). Vocabulary frequencies per grade level have also been computed but not shown here.

Toreador is a tool which runs independently of Read-X and it's designed to predict unknown vocabulary for specific reader and grade profiles currently

specified by the user. A screenshot of Toreador is shown in Figure (2). The interface shows two tabs labeled "Enter text here" and "Read text here". The "Enter text here" tab allows the user to customize vocabulary difficulty predictions by selecting the desired grade or theme.⁴ Then, text can be copied from another source and pasted in the window of the tool. The tool will analyze the text and in a few seconds return the results for the analysis in the tab labeled "Read text here", shown in Figure (3). Toreador checks the vocabulary frequency of the words in the pasted text and returns the text highlighted with the words that do not rank high in the vocabulary frequency index for the chosen categories (grade or theme). The highlighted words are clickable. When they are clicked, they entry information from WordNet appears on the right panel. The system has not been evaluated yet so some tuning will be required to determine the optimal cut-off frequency point for highlighting words. An option is also available to deactivate highlights for ease of read or reading for global meaning. Words that the system has

⁴The screenshot in Figure (2) shows an earlier version of the tool where only three thematic categories were available.

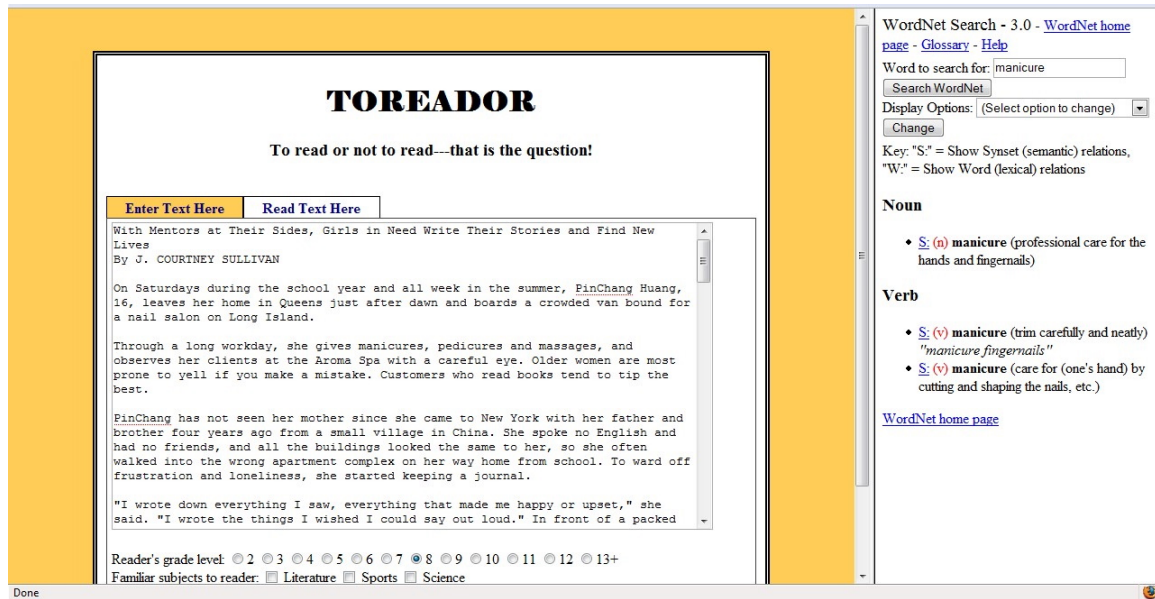


Figure 2: Text analysis of vocabulary difficulty

Arts	Career and Business	Literature	Philosophy	Science	Social Studies	Sports, Health	Technology
Word Freq	Word Freq	Word Freq	Word Freq	Word Freq	Word Freq	Word Freq	Word Freq
musical 166	product 257	seemed 1398	argument 174	trees 831	behavior 258	players 508	software 584
leonardo 166	income 205	myself 1257	knowledge 158	bacteria 641	states 247	league 443	computer 432
instrument 155	market 194	friend 1255	augustine 148	used 560	psychoanalytic 222	player 435	site 333
horn 149	price 182	looked 1231	belief 141	growth 486	social 198	soccer 396	video 308
banjo 128	cash 178	things 1153	memory 130	acid 476	clemency 167	football 359	games 303
american 122	analysis 171	caesar 1059	truth 130	years 472	psychology 157	games 320	used 220
used 119	resources 165	going 1051	logic 129	alfalfa 386	psychotherapy 147	teams 292	systems 200
nature 111	positioning 164	having 1050	things 125	crop 368	united 132	national 273	programming 174
artist 104	used 153	asked 1023	existence 115	species 341	society 131	years 263	using 172
wright 98	sales 151	indeed 995	informal 113	acre 332	court 113	season 224	engineering 170

Table 3: 10 top most frequent words per thematic category.

not seen before, count as unknown and can be erroneously highlighted (for example, the verb “give” in the screenshot example). We are currently running evaluation studies with a group of volunteers. While we recognize that the readability formulas currently implemented in Read-X are inadequate measures of expected reading difficulty, Toreador is not designed as an improvement over Read-X but as a component measuring expected vocabulary difficulty. Other factors contributing to reading difficulty such as syntactic complexity, propositional density and rhetorical structure will be modeled separately in the future.

8 Summary and future work

In this paper we presented preliminary versions of two tools developed to assist struggling readers identify text that is at the desired level of reading diffi-

culty while at the same time interesting and relevant to their interests. Read-X is, to our knowledge, the first system designed to locate, classify and analyze reading difficulty of web text in real time, i.e., performing the web search and text analysis in seconds. Toreador analyzes the vocabulary of given text and predicts which words are likely to be difficult for the reader. The contribution of Toreador is that its predictions are based on vocabulary frequencies calculated per thematic area and are different depending on the reader’s prior familiarity with the thematic areas.

We emphasize the shortcomings of the existing readability formulas, currently implemented in Read-X, and the need to develop more sophisticated measures of reading difficulty. We recognize that perceived difficulty is the result of many factors, which need to be analyzed and modeled separately.

TOREADOR

To read or not to read—that is the question!

Enter Text Here Read Text Here

With **Mentors** at their **Sides**, Girls in Need Write Their Stories and Find **New Lives** By J. COURTNEY SULLIVAN On Saturdays **during** the school **year** and all week in the summer, **PinChang Huang**, 16, leaves her home in Queens just after dawn and boards a crowded van **bound** for a nail **salon** on Long **Island**. **Through** a long workday, she **gives manicures, pedicures** and **massages**, and observes her clients at the **Aroma Spa** with a careful **eye**. **Older women** are **most** prone to yell if you make a mistake. **Customers** who read books **tend** to tip the best. **PinChang** has not seen her mother **since** she came to **New York** with her father and brother **four** years ago from a small village in China. She spoke no **English** and had no friends, and all the **buildings** looked the **same** to her, so she **often** walked into the wrong apartment **complex** on her way home from school. To ward off **frustration** and loneliness, she started **keeping** a journal. "I wrote down everything I saw, everything that made me happy or upset," she said. "I wrote the things I wished I could say out loud." In front of a packed auditorium at the **New School** in Manhattan one night **recently**, she got her chance. **PinChang** and her **mentor**, Deborah **Kolben**, a former managing editor at The Village Voice, read an essay they wrote together about **PinChang's** getting her **first manicure**. **PinChang** spoke about the **peculiar** sensation of **being** on the receiving **end** of an **exchange** that **often** makes her feel "like a slave." The reading was hosted by a nonprofit **group** that pairs high school girls from disadvantaged backgrounds who want to be **writers** with **women** who are **authors, journalists, playwrights, poets** and **editors**. The **group** produces an **anthology** of student writing each spring, and puts on **several public** readings. The readings are **often** the first chance girls get to read their own words in front of an **audience**. "I was so **nervous** when I stepped **onstage**," recalled **PinChang**, a junior at **Flushing International High School**. "I was shaking. But now I feel like I can say or do anything." **Maya Nussbaum**, 31, helped found the **group**. Girls Write Now, 10 **years** ago when she was a **senior** **majoring** in creative writing at

WordNet Search - 3.0 - WordNet home page - Glossary - Help
 Word to search for: manicure
 Search WordNet
 Display Options: (Select option to change)
 Change
 Key: "S" = Show Synset (semantic) relations, "W" = Show Word (lexical) relations
Noun
 • S: (n) manicure (professional care for the hands and fingernails)
Verb
 • S: (v) manicure (trim carefully and neatly "manicure fingernails")
 • S: (v) manicure (care for (one's hand) by cutting and shaping the nails, etc.)
[WordNet home page](#)

Figure 3: Text analysis of vocabulary difficulty

Our goal in this research project is not to provide a single readability score. Instead, we aim at building models for multiple factors and provide individual evaluation for each, e.g., measures of syntactic complexity, ambiguity, propositional density, vocabulary difficulty, required amount of inference to identify discourse relations and prior knowledge of the reader.

In future work, several studies are needed. To achieve satisfactory performance for the fine grained thematic categories, we are collecting more data. We also plan to run the subcategories classification not as an independent classification task but as subclassification task on supercategories. We expect that the accuracy of the classifier will improve but we also expect that for very fine thematic distinctions alternative approaches may be required (e.g., give special weights for key vocabulary that will distinguish between sports subthemes) or develop new classification features beyond statistical analysis of word distributions.

More sophisticated textual, semantic and discourse organization features need to be explored which will reflect the perceived coherence of the text beyond the choice of words and sentence level structure. The recently released Penn Discourse Tree-

bank 2.0 (Prasad et al., 2008)⁵ is a rich source with annotations of explicit and implicit discourse connectives and semantic labels which can be used to identify useful discourse features. Finally, more sophisticated models are needed of reader profiles and how they impact the perceived reading difficulty of the text.

9 Acknowledgments

We are grateful to Mark Dredze for his help running MIRA and Ani Nenkoca for useful discussions on readability. We thank the CLUNCH group at the Computer and Information Science department at the University of Pennsylvania and two reviewers for their very useful feedback. This work is partially funded by the GAPSA/Provosts Award for Interdisciplinary Innovation to Audrey Troutt, University of Pennsylvania.

References

- Jonathan Anderson. 1983. Lix and rix: Variations of a little-known readability index. *Journal of Reading*, 26(6):490–496.
- M Coleman and T. Liau. 1975. A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, 60:283–284.

⁵Project site, <http://www.seas.upenn.edu/~pdtb>

- K. Collins-Thompson and J. Callan. 2004. Information retrieval for language tutoring: An overview of the REAP project. In *Proceedings of the Twenty-Seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (poster description)*.
- Koby Crammer, Mark Dredze, John Blitzer, and Fernando Pereira. 2008. Batch performance for an online price. In *The NIPS 2007 Workshop on Efficient Machine Learning*.
- Lou Denti. 2004. Introduction: Pointing the way: Teaching reading to struggling readers at the secondary level. *Reading and Writing Quarterly*, 20:109–112.
- Ted Hasselbring and Laura Goin. 2004. Literacy instruction for older struggling readers: What is the role of technology? *Reading and Writing Quarterly*, 20:123–144.
- M. Heilman, K. Collins-Thompson, J. Callan, and M. Eskenazi. 2006. Classroom success of an intelligent tutoring system for lexical practice and reading comprehension. In *Proceedings of the Ninth International Conference on Spoken Language Processing*.
- M. Heilman, K. Collins-Thompson, J. Callan, and M. Eskenazi. 2007. Combining lexical and grammatical features to improve readability measures for first and second language texts. In *Proceedings of the Human Language Technology Conference, Rochester, NY*.
- Derrick Higgins, Jill Burstein, Daniel Marcu, and Claudia Gentile. 2004. Evaluating multiple aspects of coherence in student essays. In *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL 2004)*.
- Eleni Miltsakaki and Karen Kukich. 2004. Evaluation of text coherence for electronic essay scoring systems. *Natural Language Engineering*, 10(1).
- Sarah Petersen and Mari Ostendorf. 2006. Assessing the reading level of web pages. In *Proceedings of Interspeech 2006 (poster)*, pages 833–836.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The penn discourse treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- T. D. Snyder, A.G. Tan, and C.M. Hoffman. 2006. Digest of education statistics 2005 (nces 2006-030). In *U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office*.
- John Strucker, Yamamoto Kentaro, and Irwin Kirsch. 2007. The relationship of the component skills of reading to oral performance: Tipping points and five classes of adult literacy learners. In *NCSALL Reports* 29. Boston: National Center for the Study of Adult Learning and Literacy (NCSALL).
- Thomas Weinstein and Herbert J. Walberg. 1993. Practical literacy of young adults: educational antecedents and influences. *Journal of Research in Reading*, 16(1):3–19.