Literatin: Beyond Awareness of Readability in Terms and Conditions

Stuart Moran
Stuart.Moran@nottingham.ac.uk

Ewa Luger
Ewa.Luger@nottingham.ac.uk

Tom Rodden
Tom.Rodden@nottingham.ac.uk

Mixed Reality Lab
University of Nottingham
Nottingham, UK

Copyright is held by the owner/author(s). Publication rights licensed to ACM. 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.
UbiComp '14, September 13 - 17 2014, Seattle, WA, USA
Copyright 2014 ACM 978-1-4503-3047-3/14/09…$15.00.
http://dx.doi.org/10.1145/2638728.2641684

Abstract
Terms and Conditions (T&Cs) are frequently unread as a consequence of their complexity and length. Readability formulas are used to objectively measure this complexity, but ironically their outputs are also unreadable to many. This motivated the development of a chrome extension called Literatin that compares the complexity of popular fictional literature to T&Cs in order sensitise people to their complexity. In this paper we discuss whether this has been achieved, and outline plans to further develop the extension.

Author Keywords
Consent; Chrome; Extension; Readability; SMOG

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Through our lives we make use of many different services and devices, most of which require us to read and agree to a series of Terms and Conditions (T&Cs). Unfortunately, due to the necessity for companies to protect themselves and their products, a great deal of complex legal text ('legalease') is often found in T&Cs. As a result of this it is common that users of the service/device rarely read or fully comprehend what

they are agreeing to. This lack of comprehension/ readability is problematic for those seeking to give *informed* consent as this requires the *comprehension* of the consent-giver. Our prior work made the case that accessibility in HCI should include consideration of the comprehensibility of documents intended to support the consent process, such as T&Cs and privacy policies. Readability is a primary indicator of comprehensibility. By comparing the readability of documents to the reading age of the wider English population we were able to show the proportion of people excluded from the consent process; over half.

However, in order to make such comparisons, we needed to surface the readability of such documents. One way to objectively measure such a lack of readability in text is by using so called 'readability formulas' which are established on the basis of sentence length, word counts and polysyllabic words [1]. The ability to assess the readability of T&Cs could help users better appreciate whether they are informed or not. However, the irony inherent in this process is the 'unreadability' of the results from readability formulas. This motivated us to explore more understandable options for engaging users with the readability of texts. Our solution was to ground the results in users' own knowledge and experience, to create a more effective comparison point. This led to the development Literatin [3], a chrome extension designed to help users better understand the complexity of texts by presenting them with a comparison to literature they are likely to have read; both popular and classical. This paper provides a reflection upon progress since the development of Literatin. We describe the design and continued implementation of the extension, examine its subsequent impact since deployment and, given its ongoing relevance, discuss ways of developing the application further as a more stable and responsive instrument.

Design

The problem with readability formulas is that the statistics they produce are often difficult for the layperson to contextualize. One solution to this, as outlined by Walsh and Volsko [5], is to graphically represent the results, anchoring the visualization to something users can relate to. Taking this idea forward, we sought to find a graphical reference point with which users could make judgments about the complexity of texts based partially on their own experiences. Given the focus on the readability of text, it seemed apt to make use of modern fictional literature that users are likely to have read personally or in school. By situating these books alongside each other, on a scale of complexity, their familiarity acts as a suitable reference point, eliminating need for statistics whilst also minimizing interpretation time. Research shows that users spend less than a second reading T&Cs and so our solution needed to be immediate. We chose to create this tool as an extension for a web browser so that comparison could take place in situ, as users come across terms and conditions through their routine endeavors.

Implementation

The open source extension was developed for the Chrome browser, and can be downloaded by searching for "Literatin" in the Google Play Store or visiting http://literatin.wordpress.com/. The extension uses an adaptation of the frequently used SMOG (Simple

Measure of Gobbledygook) formula developed by McLaughlin [4] to asses readability:

$$grade = 1.043 \sqrt{number\ of\ polysyllables \times \frac{30}{number\ of\ sentences}} + 8.1291$$

In our adaptation, in place of the 8 applied at the end of the above formula, we apply 3 in order to achieve the readability level of a text, rather than the reading age required. This adaptation was developed by literacy experts to more accurately reflect the adult context and, as such, is the only readability formula comparable to adult literacy.

Once the text has been cleaned, including the removal of incorrect spaces, punctuations lines and characters, complex words are calculated. This is achieved by counting three or more syllables in a word (polysyllabic). A syllable is determined based on the presence of a number of pre-determined vowel combinations.

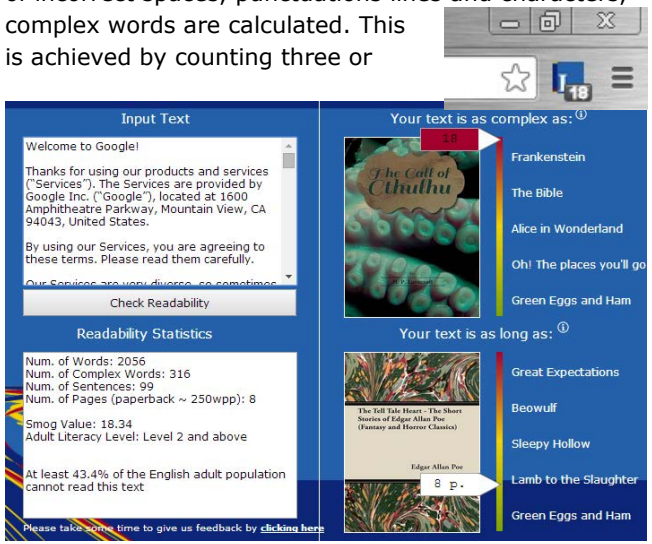

Figure 1: Literatin: A chrome extension to asses readability of text on webpage, including initial notification value (top right).

The extension works by automatically capturing the text on a webpage as it loads, and calculates the readability of the text using SMOG (locally on the user's machine). This is initially displayed as a small unobtrusive notification, to give those familiar with the SMOG values a fast means of judging page complexity. For more information on the readability, users may click the Literatin logo which loads a new screen (as seen in Figure 1).

Initially results are presented in text format, including information on Number of Words, Number of Complex Words, Length in Pages and the SMOG Value. The stats are also cross checked with government figures showing the reading age of the English population to illustrate what proportion may not be able to read the text. Using a database of 50 books, which were tested for readability, visual comparisons can be made between the text and those books. Other modes of using the app include the ability to highlight specific text on a page of interest and also enter data directly into Literatin.

Impact and Engagement

Since the extension was made publicly available in early 2013, it has been uniquely downloaded over 1,200 times across multiple disciplines; Law, Journalism, Finance, Oceanography, Media & Entertainment, Psychology, Creative writing, Engineering, Programming, Sales, Information Science and Politics. In one notable instance, a user used the tool to assess the terms and conditions of betting

websites, reporting the results to their community [2]. The work is best summarized by another user who states: "*I am not a literacy expert but I still think that this seems like a useful tool*" [25], reaffirming the extension as a tool for assessment of readability.

In terms of media coverage, the extension was reported on by major British newspapers the Times [6] and the Mail [7] in addition to Forbes. The work has also been reported on in English [8-14], Czech [15], Finnish [16-17], Spanish [18-19], Chinese [20], German [21], Afrikaans [22] and Turkish [23] news outlets, and appeared on a Chinese news channel [24].

The focus of these articles was on how Literatin helped highlight the ways complex terms and conditions unacceptably lead to the alienation of wide groups of people. One of the reasons the plug-in was so successful is the novel way it represented the complexity of text, such that a user can very easily relate to it from their own experiences. For many, the results were seen to be comical, with some terms and conditions found to require high levels of readability equal to or greater than notoriously complex text such as 'War and Peace'. However, after the amusement subsides, a serious realization emerges that documents that should be understandable by all, are not. In summary, the plug-in captures users attention by providing amusing social commentary on the readability of terms and conditions (in a way that users can comprehend), which in turn reveals to them a more serious and wider message about accessibility and readability. This more serious message also found its way to the House of Commons Science and Technology Committee in June 2014.[1]

The global news coverage demonstrates successful engagement with members of the international community, and provides suitable evidence that the tool has helped raise awareness of the issues of complexity in T&Cs. However, we believe there is more to be gained from this type of technology than simply an awareness raising exercise.

Beyond Raising Awareness

In this section, we outline a series of objectives for future development of this work that will not only continue to contribute to awareness of these critical issues, but also extend to provide a means with which to actively help writers of complex documents to make them more readable.

Granularity of Assessment

The values generated by the extension are exclusive to the SMOG formula, while others are widely available. We aim to introduce additional measures of readability so as to provide different perspectives and allow for comparison between methods and to allow comparisons against school-age. We also plan to include the ability to filter by certain polysyllabic words from the analysis. This would allow the extension to be more widely used in fields with common complex words, such as healthcare and fitness.

1 Reference to Literatin at 16:33:30

http://www.parliamentlive.tv/Main/Player.aspx?meetingId=15596

[Accessed 30/06/2014]

Relevance to the User
Currently the extension includes a set of 50 books, which is limited in cultural variety. To help further the relevance of the extension to users, we aim to not only include a wider variety of literature, but to also allow users to personalize the extension using their own books or other texts relevant to them. Much of the statistics are also focused on a British system of assessing readability, and this will be expanded to include a more appropriate metric depending on the location of the user. We also aim to expand the extension to other browsers, including internet explorer, Fire Fox and safari to widen the accessibility of the extension to others.

Creating Change through Feedback
One of the major additions to the extension is the need to give feedback to users on how the input text might be made less complicated. We aim to highlight a range of complex words throughout the text with colours which show the degrees of complexity (e.g. red for most complex, green for least). On top of this, a system will suggest replacements with less complicated words, and also demonstrate how much this will reduce the overall complexity of the text by. We will also introduce a series of analytic tools that allow us to understand how the extension is used, to make improvements.

Example Application Scenarios
Most text documents are written in a word processor, and so our main goal for future development of the extension will be to support the writing process as it occurs. We aim to create an extension for word processors that will perform text analysis on the fly, providing feedback to users as they type. We also hope to encourage a standard to be developed as a means of cross comparing readability internationally.

Conclusions
Terms and Conditions are frequently not read because they are long and complex. This complexity arguably prevents people from giving truly informed consent. This lack of readability can be objectively measured using readability formulas, which themselves are unreadable. As such, we developed a chrome extension to help raise awareness of the complexity of terms and conditions.

In this paper, we report on the success of this endeavor with wide international public adoption of the extension, and outline a series of future changes that will shift the focus of the extension from raising awareness, to creating change. This will be achieved by giving direct feedback to the creators of terms and conditions during the writing process.

Acknowledgements
This work was supported by EPSRC grant EP/I011587/1.

References
[1] Bailin, A. and Grafstein, A. The linguistic Assumptions Underlying Readability Formulae: A Critique. Language and Communication 21, 3 (2001), 285–301.

[2] Høgenhaven, T. Bookmaker Terms & Conditions: As Hard To Read As Nietzsche. http://www.bettingexpert.com/blog/terms-conditions-study, 2013.

[3] Luger, E., Moran, S., and Rodden, T. Consent for All: Revealing the Hidden Complexity of Terms and Conditions. Proceedings of the SIGCHI Conference on

Human Factors in Computing Systems (CHI), (2013), 2687 – 2696.

[4] McLaughlin, G.H. SMOG Grading; A New Readability Formula. Journal of Reading 12, 8 (1969), 639–646.

[5] Walsh, T. and Volsko, T.A. Readability Assessment of Internet-Based Consumer Health Information. Respiratory Care 53, 10 (2008), 1310–1315.

[6] http://www.thetimes.co.uk/tto/science/article3898546.exe *[Accessed on 30/06/2014]*

[7] http://www.dailymail.co.uk/news/article-2467166/Googles-terms-conditions-beat-classics-Beowulf-War-Peace-named-epic-read-time-say-scientists.html *[Accessed on 30/06/2014]*

[8] http://newint.org/blog/2013/10/23/googles-terms-unreadable-literatin/ *[Accessed on 30/06/2014]*

[9] https://ipdraughts.wordpress.com/2013/10/19/googles-terms-more-complex-than-beowulf/ *[Accessed on 30/06/2014]*

[10] http://theconversation.com/googles-terms-and-conditions-are-less-readable-than-beowulf-19215 *[Accessed on 30/06/2014]*

[11] http://www.sourcewire.com/news/77683/bookie-t-cs-as-hard-to-read-as-nietzsche- *[Accessed on 30/06/2014]*

[12] http://www.forbes.com/sites/timworstall/2013/10/20/googles-tcs-should-be-more-50-shades-of-grey-and-less-beowulf/ *[Accessed on 30/06/2014]*

[13] http://qz.com/136688/googles-new-terms-of-service-are-harder-to-read-than-beowulf/ *[Accessed on 30/06/2014]*

[14] http://www.bettingexpert.com/blog/terms-conditions-study *[Accessed on 30/06/2014]*

[15] http://technet.idnes.cz/podminky-google-0j4-/sw_internet.aspx?c=A131021_150903_sw_internet_mla *[Accessed on 30/06/2014]*

[16] http://www.mbnet.fi/artikkelit/sota_ja_rauha_helpompi_kuin_googlen_kayttoehdot *[Accessed on 30/06/2014]*

[17] http://www.marmai.fi/uutiset/onko+homeroksen+ilias+mielestasi+vaikealukuinen+kokeile+lukea+googlen+ja+facebookin+sopimusehtoja/a2210292 *[Accessed on 30/06/2014]*

[18] http://alt1040.com/2013/10/condiciones-legales-google *[Accessed on 30/06/2014]*

[19] http://noticias.terra.com.pe/tecnologia/internet/condiciones-legales-de-google-dificiles-de-entender,f8cf37ae077d1410VgnVCM3000009af154d0RCRD.html *[Accessed on 30/06/2014]*

[20] http://paper.wenweipo.com/2013/10/20/YO1310200011.htm *[Accessed on 30/06/2014]*

[21] http://www.heise.de/tp/artikel/40/40138/1.html *[Accessed on 30/06/2014]*

[22] http://www.n24.de/n24/Nachrichten/n24-netzreporter/d/3692136/ein-schinken-wie--krieg-und-frieden-.html *[Accessed on 30/06/2014]*

[23] http://www.taraf.com.tr/haber-google-savas-ve-baris-i-yendi-137417/ *[Accessed on 30/06/2014]*

[24] http://www.ntdtv.com/xtr/gb/2013/10/19/atext986872.html *[Accessed on 30/06/2014]*

[25] http://wecdsbit.blogspot.co.uk/2014/02/how-readable-is-your-text.html *[Accessed on 30/06/2014]*