Gmail **Joao Palotti <joaopalotti@gmail.com>**

WWW 2018 notification for paper 341

WWW 2018 <www2018@easychair.org> Thu, Dec 21, 2017 at 8:29 PM
To: Joao Palotti <joaopalotti@gmail.com>

Dear Joao:

We regret to inform you that your submission was not selected for presentation at the research track of “The Web Conference 2018” (WWW 2018):

Paper: 341
Title: A Study of Web Page Understandability for Consumer Health Search

This year we received 1155 valid submissions, 19% more than last year. We were able to accept only 171 of these, representing an acceptance rate of about 14.8%. Due to the limited number of available slots in the conference schedule, we unfortunately had to make difficult decisions, and many worthy submissions were not accepted to the research track.

The program committee worked very hard to thoroughly review all the submitted papers and to provide action points to improve your paper. All papers were reviewed by at least three program committee members (and many of the papers were reviewed by four or more), by track chairs, and in some tracks also by a senior PC member. The discussions were extensive, and a large number of submissions was discussed among the track chairs and the program chairs during an in-person PC meeting.

Please consider submitting your work to one of these other venues associated with WWW 2018. The majority of these venues have their submission deadlines set on January 5 and 10, 2018 (but please double check the individual websites below):

Posters: https://www2018.thewebconf.org/call-for-papers/posters-cfp/
Demos: https://www2018.thewebconf.org/call-for-papers/demos-track-cfp/
Workshops: https://www2018.thewebconf.org/call-for-papers/workshops-cfp/

Alternate tracks on

Journalism, Misinformation, and Fact Checking: https://www2018.thewebconf.org/call-for-papers/misinformation-cfp/
The BIG Web (formerly BigData Innovators Gathering): https://www2018.thewebconf.org/call-for-papers/the-big-web-cfp/

We very much hope you will be able to attend WWW 2018 in April in Lyon, France. Further information about the conference can be found at https://www2018.thewebconf.org/.

Best regards,
WWW 2018 PC Chairs
Panos Ipeirotis and Mounia Lalmas

---------- REVIEW 1 ----------
PAPER: 341
TITLE: A Study of Web Page Understandability for Consumer Health Search
AUTHORS: Joao Palotti, Guido Zuccon and Allan Hanbury

Audience: 3 (Yes, to a large group of people)
Overall score: 1 (weak accept)

---------- Strength ----------
The paper provides a very thorough set of experiments exploring interactions between understandability and retrieved relevant web pages in the health context.

---------- Weakness ----------
The Tables and Figures are very difficult to read, and there are too many abbreviations and non-mnemonic names (e.g. "Combo X") to keep track of. Some summary plots/more synthesis would assist that.

---------- Summary and review comments ----------
The authors have experimented first with estimating the understandability of web pages, including features that address lexical complexity, based on CLEF data that includes understandability assessments; correlations between automatically computed estimates and human estimates are provided. The authors also explore the impact of variation in pre-processing steps on the understandability metrics. Finally, the information retrieval context is directly considered, and results are presented using metrics that combine relevance and understandability.

The authors have incorporated a several large external resources such as Medical Reddit and Wikipedia, in addition to the published literature collected for TREC Clinical Decision Support tracks, to establish word familiarity, and to provide estimates of understandability used to train understandability classifiers, based on assumed formality of the different resources (producing a "silver standard" labeling that does not necessarily correspond to human judgements for individual documents).

The understandability classification models are constructed using a LSA-reduced space of 10 dimensions; the authors state that this was empirically set based on document word counts. More detailed experimentation with the choices of the feature representation and methods for these classifiers is warranted. Some sense of how well the classifiers are doing against the "silver standard", e.g. in a cross-fold validation framework), would be interesting. I do realize that this is not the main point of the paper, but given that these models work quite well, relatively, it would be worthwhile to expand on this point.

As a small comment: MeSH entities are not the complete Expert Medical Vocabulary in UMLS/MetaMap; why is MeSH singled out for this?

Performance on CLEF 2016 generally seems to be worse than for CLEF 2015 (certainly for the regression/classification methods) -- why is this the case?

I wasn't fully clear on how to interpret the box plots in Figure 2 -- the text indicates that this represents "variability in terms of correlation with human assessors" but what is the variability over given that each line on the x axis corresponds to a given pre-processing method + understandability estimator combination? Isn't a single ranking computed for each of these combinations over one data set? Perhaps I missed something.

There are some minor typos:
elderlies -> elderly
natural language and language -> natural language processing and language
correlations were archived by -> correlations were achieved by

---------- REVIEW 2 ----------
PAPER: 341
TITLE: A Study of Web Page Understandability for Consumer Health Search
AUTHORS: Joao Palotti, Guido Zuccon and Allan Hanbury

Audience: 2 (Yes, but only to a small group of people)
Overall score: -1 (weak reject)

---------- Strength ----------
the work brings attention to estimation of understandability of health-related web-pages. Promises to make the data and code publicly available

---------- Weakness ----------
: the issues of understandability are well known and much of what the authors have to say has been told before.

---------- Summary and review comments ----------
This paper explores understandability and topical relevance of some online consumer-health resources. The work is based on the data from the CLEF 2015 and 2016 eHealth collections. The authors confirm that the standard readability metrics are not a good approximation of consumers' potential to understand health-related resources and that preprocessing of the html pages interferes with the automated approaches to estimate understandability of these pages. The authors also attempt to re-rank pages with respect to understandability while maintaining topical relevance of the results. Although the paper is well-written and the study is relatively thorough, I have to disagree with the conclusion that "This paper makes a clear contribution to improving search engines tailored to consumer health search because it thoroughly investigates promises and pitfalls of understandability estimations and their integration into retrieval methods " The contribution is incremental at best!
and no clear paths to improvements are demonstrated.

---------- REVIEW 3 ----------
PAPER: 341
TITLE: A Study of Web Page Understandability for Consumer Health Search
AUTHORS: Joao Palotti, Guido Zuccon and Allan Hanbury

Audience: 2 (Yes, but only to a small group of people)
Overall score: 1 (weak accept)

---------- Strength ----------
Two main strengths: 1) the rationale for the paper was well developed and 2) the methods were detailed and comprehensive.

---------- Weakness ----------
I understand that this was a methods-focused paper and they were detailed in their approach; however, I was left wanting more of an implications or "so what" connection, more than the last paragraph relating back to figure 1. The three points described in the conclusion are good methodological contributions, and I would like to see these tied back to practical implications.

---------- Summary and review comments ----------
The purpose of this paper was to propose and examine methods for the estimate of the understandability of health information in Web pages. The authors provide a strong rationale for why this paper would add to the literature and provide detailed analysis of their methodological approaches. The use of Medical Reddit, Medical English Wikipedia, and PubMed Central helps strengthen the paper by providing a clear path for application of findings; however, this could have been better developed in the paper. For example, findings from the empirical experiments suggest that machine learning methods are useful in estimating the understandability of Web pages - how then does this factor back to the website used in the analysis? How does this 'look' or 'play out' in terms of how the findings can be applied to current practices and standards and assist consumers of health information. Overall, an interesting paper.

---------- REVIEW 4 ----------
PAPER: 341
TITLE: A Study of Web Page Understandability for Consumer Health Search
AUTHORS: Joao Palotti, Guido Zuccon and Allan Hanbury

Audience: 2 (Yes, but only to a small group of people)
Overall score: -2 (reject)

---------- Strength ----------
This topic of the paper is significant. Beyond query relevance, understandability of a health page online is an important criteria to include in online searches. The authors show a thorough set of experiments on a benchmark dataset used in community challenges in 2015 and 2016.

---------- Weakness ----------
The primary weakness of the paper is its novelty from a methods and from a task standpoint. While the authors are rigorous in their experiments, they do not contribute any novel approach to characterizing understandability/readability of a health page. The finding that machine learning techniques fare better than readability metrics has been known for a while when it comes to health texts as well as texts in the general news domain.

---------- Summary and review comments ----------
In this paper, the authors conduct a series of thorough experiments to estimate understandability of health Web pages and incorporating these estimates as part of a re-ranking within a search architecture. For their experiments, the authors use the datasets from the CLEF 2015 and 2016 challenges, which consist of query, web pages, and associated undertandability ratings. The primary concern with the paper is its lack of novelty. Casting the task of understandability as a classification task, and engineering features has been the subject of a considerable amount of work across communities of NLP, IR, and Informatics. There is an opportunity for novel work here, and thinking of maybe more "actionable" assessment of understandability (e.g., would text simplification help?).

- In the natural language features, the only features included are part of

speech classes, sentiment polarity, and ratio of words found in English vocabulary. What is the motivation behind using sentiment polarity? Further, work in readability/understandability in the general English domain has established that syntactic features and discourse features are particularly helpful for assessing understandability. Why were these not considered? In addition to the work cited in the manuscript, which refers to use of syntax, see these two papers about use of discourse features:
* Feng, Lijun, et al. "A comparison of features for automatic readability assessment." Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010.
* Barzilay, Regina, and Mirella Lapata. "Modeling local coherence: An entity-based approach." Computational Linguistics 34.1 (2008): 1-34.

- The word frequency features make sense, and has also been investigated in the context of readability of health texts, see for instance
* Elhadad, Noemie. "Comprehending technical texts: Predicting and defining unfamiliar terms." AMIA annual symposium proceedings. Vol. 2006. American Medical Informatics Association, 2006.
* Wu, Danny TY, et al. "Assessing the readability of ClinicalTrials. gov." Journal of the American Medical Informatics Association 23.2 (2015): 269-275.

Another weakness of the paper concerns the understandability assessments obtained as part of the CLEF datasets. Can the authors explain better how valid the assessments are and how generalizable they would be? Another way to make the work more innovative would be to study the generalizability of the understandability assessment across health consumers with different levels of health literacy for instance.

Finally, this is a small point, the Introduction motivates the need for understandability assessment by talking about misleading information. The two concepts are independent of each other. An easy-to-read document may or may not have any misleading information, and knowledge about one concept would not help towards knowing the others.

---------- METAREVIEW ----------
PAPER: 341
TITLE: A Study of Web Page Understandability for Consumer Health Search

RECOMMENDATION: reject

This paper examines methods for estimating the understandability of health information in Web pages, and applying the estimated understandability during retrieval. The reviewers commended the paper on the thoroughness of the paper. They were concerned by the novelty of the paper, especially given that the main insights of the paper have been shown in prior work.