Effect of the Named Entity Recognition and Sliding Window on the HONcode Automated Detection of HONcode Criteria for Mass Health Online Content

Celia Boyer[1], Ljiljana Dolamic[1], Patrick Ruch[2] and Gilles Falquet[3]
[1]*Health On the Net Foundation, Chemin du Petit-Bel-Air 2, Chłne Bourg, Switzerland*
[2]*HES-SO Geneva and SIB Swiss Institute of Bioinformatics, Geneva, Switzerland*
[3]*Faculty of Economics and Social Sciences, University of Geneva, Geneva, Switzerland*

Keywords: HONCODE, Automated Detection, Manual Detection, Machine Learning, Named Entity Recognition.

Abstract: The Health On the Net's Foundation (HON) Code of Conduct, HONcode, is the oldest and the most used ethical and trustworthy code for medical and health related information available on the Internet. Until recently, websites voluntarily applying for the HONcode seal were evaluated manually by an expert medical team according to 8 principles, referred to as criteria, and associated published guidelines. In the scope of the European project Kconnect, HON is developing an automated system to identify the 8 HONcode criteria within health webpages. When the research on the development of such a system evolved from simple algorithmic testing to a real full-content setting, it revealed a number of issues. The preceding study consisted in taking a set of 27 health-related websites and having them assessed for their compliance to each of the 8 HONcode criterion, first manually by senior HONcode experts, and then through supervised machine learning by the automated system. The results showed discrepancies mainly for two criteria: "submerged content" under the *Complementarity* criterion and "extremely low recall" under the *Date Attribution* criterion. In this article, the authors investigate different approaches to solve the problems related to each of these criteria, namely a customized Named Entity Recognition Model instead of a machine learning component for Date Attribution, and a sliding window instead of the whole document as a unit of detection for *Complementarity*. The results obtained show that the newly adapted automated system greatly improves accuracy: 74% vs. 41% for the *Date Attribution* criterion and 74% vs. 22% for the *Complementarity* criterion.

1 INTRODUCTION

Despite the abundance of online health content, the issue lies in its reliability. This problem is particularly acute in the medical information field, which directly involves public health (van Straten et al., 2008; Humphrey, 2009). Efforts have been taken to automatically label online health pages in line with the quality of the information they provide (Aphinyanaphongs et al., 2005; Griffiths et al., 2005).

The Health on the Net Foundation established the HON Code of Conduct in 1996 (Boyer et al., 1999) with a consensus of health information editors in order to have common good practice criteria for online health information. The aim of the HONcode (summary in Table 1) is to guide Internet users and patients towards trustworthy medical information by certifying health-related websites offer content that respects a defined set of criteria. This quality label (logo or HONcode seal) is displayed on a health website to prove the provider is committed to implementing or adhering to the HONcode. It can only be boasted after submission of a formal application and approval from HON. The website is reviewed on a regular basis, and users can report misuse of the label if need be (http://services.hon.ch/Contact/contact.pl).

Considering the HONcode certification is a voluntary process performed only upon request, a website can be completely reliable and respect the HONcode criteria without being certified. Indeed, while the Internet contains thousands of health-related websites, only 8,000 of them belonged to the HONcode certified website community in 2014. This proves a large majority of health websites do not display the HONcode seal. Information seekers have found a trick to identify HONcode certified websites: they add the word "HONcode" to the health terms they put

Boyer, C., Dolamic, L., Ruch, P. and Falquet, G.
Effect of the Named Entity Recognition and Sliding Window on the HONcode Automated Detection of HONcode Criteria for Mass Health Online Content.
DOI: 10.5220/0005644301510158
In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) - Volume 5: HEALTHINF*, pages 151-158
ISBN: 978-989-758-170-0

Table 1: The eight HON Code of Conduct (HONcode) criteria for medical and health websites. http://www.hon.ch/Conduct.html.

	Criteria name	Detail
HC1	Authoritative	Indicate the qualifications of the authors
HC2	Complementarity Information	should support, not replace, the doctor-patient relationship
HC3	Privacy policy	Disclose and respect the privacy and confidentiality of personal data submitted to the site by the visitor
HC4	Attribution Reference criteria	Cite the source(s) of published information
HC4	Attribution Date	date medical and health pages
HC5	Justifiability	Site must back up claims relating to benefits and performance
HC6	Transparency	Accessible presentation, accurate email contact
HC7	Financial disclosure	Identify funding sources
HC8	Advertising policy	Clearly distinguish advertising from editorial content

in the search bar, e.g. "macular degeneration HONcode" yielded 10 results in Google and 8 in Bing: HONcode certified websites providing information on macular degeneration. HON has developed the HONcode toolbar to direct Yahoo, Google and Wikipedia users to certified websites. However, in order to use the toolbar, the user needs to be aware of the HONcode and install the toolbar on his/her browser. In 2003, Ilic and al. pointed out the limitation of specialized search-engines due to the lower volume of health information available because of the small number of indexed websites. This may also be due to the efficiency of the crawler that harvests the pages proposed by the search engine (Ilic et al., 2003).

To overcome the limitations of identifying trustworthy health websites, the European Commission under the Information and Communication Technologies (ICT) Theme of the 7th Framework Programme for Research and Technological Development funded the KHRESMOI project 2010-2014. Within this project, the Health On the Net Foundation collaborated with 11 partners to develop a multilingual multimodal search and access system for medical information and health-related documents. Its objective was to address the challenges of retrieving relevant health information among huge amounts of medical data, including general medical information available online (everyone.khresmoi.eu/everyone.khresmoi.eu/; (Boyer et al., 2014)). Today, HON is pursuing research on how the ethical principles within a health website can be identified automatically. This work is being continued by the KConnect European project and focuses on offering an integrated solution for citizens' use.

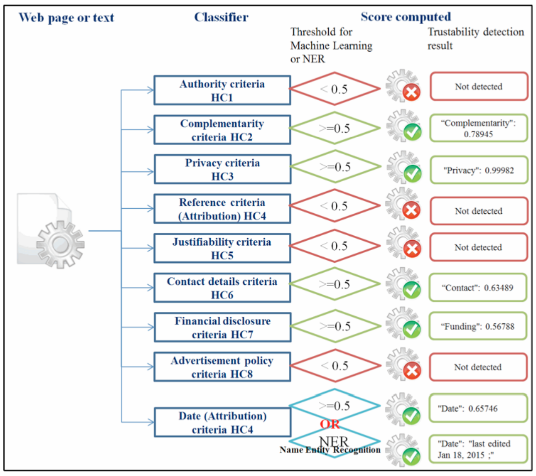
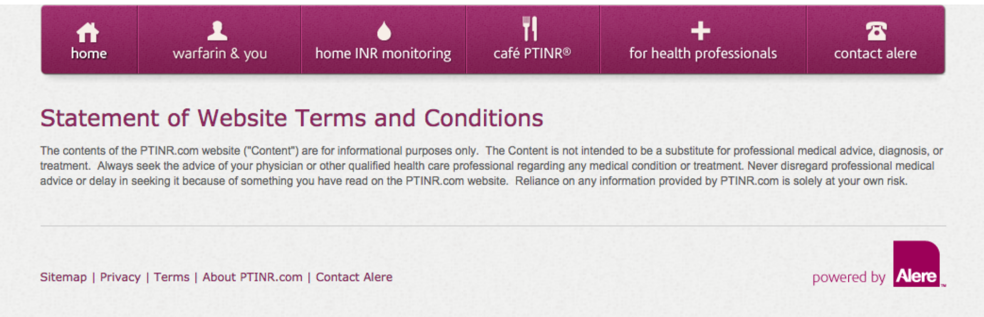

Web page or text
Classifier
Score computed
Threshold for Machine Learning or NER
Trustability detection result
Authority criteria HC1
Complementarity criteria HC2
Privacy criteria HC3
Reference criteria (Attribution) HC4
Justifiability criteria HC5
Contact details criteria HC6
Financial disclosure criteria HC7
Advertisement policy criteria HC8
Date (Attribution) criteria HC4
< 0.5
>=0.5
>=0.5
< 0.5
< 0.5
>=0.5
>=0.5
< 0.5
>=0.5
OR
NER
Name Entity Recognition
Not detected
"Complementarity": 0.78945
"Privacy": 0.99982
Not detected
Not detected
"Contact": 0.63489
"Funding": 0.56788
Not detected
"Date": 0.65746
"Date": "last edited Jan 18, 2015 ;"

Figure 1: Automated system for HONcode detection with 9 distinct classifiers.

2 MOTIVATIONS

The automated system for the detection of HONcode criteria illustrated in Figure 1 is described in detail in Boyer and Dolamic, 2014. This system consists of 9 distinct classifiers based on a machine learning framework (Williams and Calvo, 2002) for each of the HONcode criteria. The Attribution criterion is divided in two distinct parts: namely *Date Attribution* and References. The excerpts extracted by HONcode experts as a justification of the website's compliance to a given criterion are used as a learning/test collection in this system. We used the standard systematic evaluation scheme with 80% of the collection employed to train the system and 20% to assess the algorithm. The Nave Bayes machine learning algorithm with a single word tokenization and a space reduction of 70% was used for the classifier as a result of previous publication work ((Boyer and Dolamic, 2014; Boyer and Dolamic, 2015) and in this research). The evaluation of the classifier on the remaining 20% of the collection yielded good results for the *Complementarity* and *Date Attribution* criteria with a precision of 83% and 94% respectively, and a recall of 95% for both.

After extensively testing this system with the sample collection, the next logical step was to verify its performance on real life examples, thus comparing



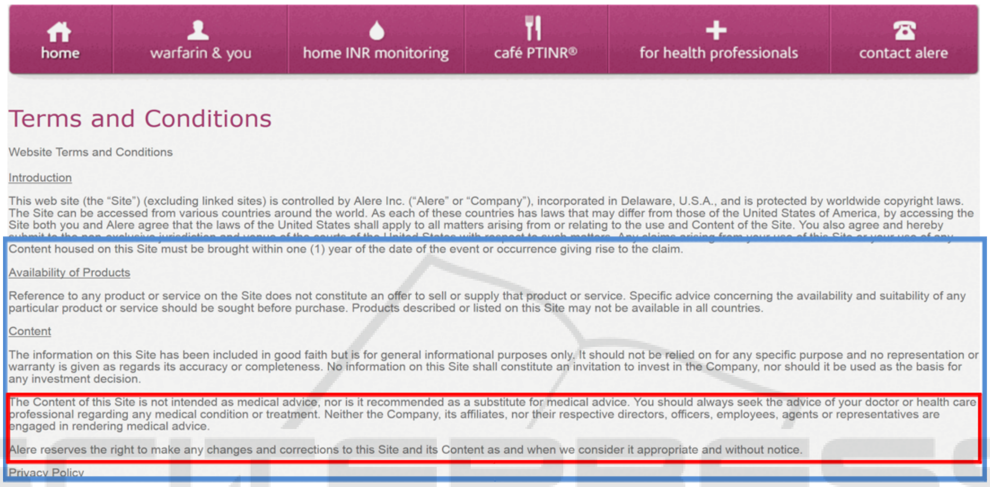
Statement of Website Terms and Conditions

The contents of the PTINR.com website ("Content") are for informational purposes only. The Content is not intended to be a substitute for professional medical advice, diagnosis, or treatment. Always seek the advice of your physician or other qualified health care professional regarding any medical condition or treatment. Never disregard professional medical advice or delay in seeking it because of something you have read on the PTINR.com website. Reliance on any information provided by PTINR.com is solely at your own risk.

Sitemap | Privacy | Terms | About PTINR.com | Contact Alere

powered by Alere

Figure 2: Complementarity criteria detected by classifier (True Positive TP) for the page http://ptinr.com/node/4478.



Terms and Conditions

Website Terms and Conditions

Introduction

This web site (the "Site") (excluding linked sites) is controlled by Alere Inc. ("Alere" or "Company"), incorporated in Delaware, U.S.A., and is protected by worldwide copyright laws. The Site can be accessed from various countries around the world. As each of these countries has laws that may differ from those of the United States of America, by accessing the Site both you and Alere agree that the laws of the United States shall apply to all matters arising from or relating to the use and Content of the Site. You also agree and hereby

Content housed on this Site must be brought within one (1) year of the date of the event or occurrence giving rise to the claim.

Availability of Products

Reference to any product or service on the Site does not constitute an offer to sell or supply that product or service. Specific advice concerning the availability and suitability of any particular product or service should be sought before purchase. Products described or listed on this Site may not be available in all countries.

Content

The information on this Site has been included in good faith but is for general informational purposes only. It should not be relied on for any specific purpose and no representation or warranty is given as regards its accuracy or completeness. No information on this Site shall constitute an invitation to invest in the Company, nor should it be used as the basis for any investment decision.

The Content of this Site is not intended as medical advice, nor is it recommended as a substitute for medical advice. You should always seek the advice of your doctor or health care professional regarding any medical condition or treatment. Neither the Company, its affiliates, nor their respective directors, officers, employees, agents or representatives are engaged in rendering medical advice.

Alere reserves the right to make any changes and corrections to this Site and its Content as and when we consider it appropriate and without notice.

Privacy Policy

Figure 3: Complementarity criteria not detected by the classifier (False Negative FN) for the page http://ptinr.com/terms-and-conditions.

it to the current manual HONcode certification process. In the report (Boyer and Dolamic, 2015), we compared the automated detection of the HONcode criteria with a manual process and we obtained mitigated results especially for the *Date Attribution* and *Complementarity* criteria. We also identified further research paths to improve our first evaluation in real life settings.

Indeed, we were able to detect two distinct problems which might be the cause of the systems poor performance on real life examples for the two criteria described above. Concerning *Date Attribution*, the system was unable to detect the information if it was displayed in numbers only (with no accompanying text), e.g. 24/08/2015. Keeping the number in the tokenization process would not be a solution as it would require that all dates be listed in the training set. Forcing the feature selection in this case might result in serious system over-fitting as all the dates could be wrongly recognized as information linked to the last update.

Another problem we identified was that the content relevant to the *Complementarity* criterion was not detected because it was submerged by the large amount of information present on the page parsed by the classifier. This issue was acknowledged as the main reason for the low recall of the *Complementarity* criterion. Figures 2 and 3 respectively illustrate the capability and the limitation of the classifier based on machine learning on different pages of the same website (e.g. http://ptinr.com/). Figure 2 illustrates the correct detection by the classifier of the criteria related to *Complementarity* as the whole page is connected to this criterion only (http://ptinr.com/node/4478). In contrast, for the http://ptinr.com/terms-and-condition page illustrated in Figure 3, the system is unable to detect this criterion even though it contains almost the same text as the previous one, highlighted by the red rectangle. It is important to notice that the http://ptinr.com/terms-and-conditions page shows a large textual content, sometimes related to different criteria. The *Complementarity* criterion corresponds to 5% of this pages content which means it is not prominent on the page and other criteria are proportionally more salient. So the classifier detects the information related to the *Privacy* and *Advertising policies* criteria, while other criteria, such as the *Complementarity*, are neglected.

We named this problem the "submerged content" issue.

The *Complementarity* criterion-related problem illustrated here was also present for other criteria such as *Date Attribution*. To deal with this issue (Boyer and Dolamic, 2015), we tested the effectiveness of using the sentence as the unit of classification. This approach was experimented for the *Privacy* and *Date Attribution* criteria on a selection of websites and proved very conclusive, especially in relation to recall for both criteria.

Indeed, the sentence unit provides very promising results, with 22 true positives detected by the system out of the 24 identified manually (92% recall and 81% precision) and 20 true positive detected by the system out of the 21 identified manually (95% recall and 74% precision) for the *Privacy* and *Date Attribution* criteria respectively. However, we determined that using the sentence as the classification unit resulted in the detection of given criteria based on a single word. This word is usually the one which has a very high probability for given criteria, e.g. "policy" for the *Privacy* criterion, or "update" for *Date Attribution*. Thus, in some cases, the Privacy criterion was detected for 99% of the website pages because there is often a link towards Privacy policy in the footer of all health websites pages.

3 METHODS

To deal with the problems described in the previous section, we adapted our system in two distinct ways. For the *Date Attribution* criterion, even if the "submerged content" problem exists, it is not prevailing. The main problem for this criterion stems from the vocabulary as it is quite specific. Table 2 provides examples of excerpts taken from the set of 2'794 training data for this criterion.

However, in certain cases the extracted content is not appropriate to justify this criterion. (See † in Table 2).

In the light of the specificity of the given vocabulary for the *Date Attribution* criterion, we opted to replace the machine learning classifier by the Named Entity Recognition (NER) tool from OpenNLP toolkit (OpenNLP, 2015). In order to obtain good results, it was necessary to program the NER model to recognise as many terms as those used to refer to the element we want to identify. Therefore, the previously mentioned corpus of 2'794 excerpts (some excerpts are shown in Table 2) was used to train the NER model to respond to the needs of Data attribution. The examples of the data given in Table 3 illustrate the NER model training collection. The 27 websites were used to identify differences between the previous machine learning method used in the article (Boyer and Dolamic, 2015) and the NER method used in this research.

Table 2: Date attribution criterion examples of excerpts.

Extract reference	Extract content
018599.HC9	This notice of privacy practices is effective February 1, 2010.
018597.HC9	Revised 5/11/06 12/21/01
018598.HC9	Date created: December 13, 2006
018608.HC9	last reviewed: 15 November 2010
018614.HC9	This page was last modified on Wednesday November 24, 2010 02:53pm
021389.HC9	22/02/2013
021387.HC9	Friday, January 04, 2013
021394.HC9†	journal list: Jan 2013 impact factor: Sep 2013 resources list: Feb 2013

Table 3: NER model training data

NER model training
This notice of privacy practices is <START:date>effective February 1, 2010 <END>.
<START:date>revised 5/11/06 12/21/01 <END>
This information was <START:date>last updated on july 17, 2007 <END>
This page was <START:date>last modified on Friday may 22, 2009 <END>02:13pm
Content last updated dynamically at <START:date>last updated sun, 28 nov 2010 <END>13:04:08 -0600
Date of first authorisation/renewal of the authorisation <START:date>12th april 2003 <END>

The classifier for the *Complementarity* criterion was efficient, scoring 83% for precision and 95% for recall when evaluated systematically on 20% of the collection. However, in the real setting, where the text to be extracted is sometimes totally drowned by other content related to other criteria, the use of the sentence as the classification unit proved to create considerable noise (False positive, e.g. pages are detected as presenting a policy information which, in fact, was only a link to the privacy policy page in the best case) (Boyer and Dolamic, 2015). In order to deal with "submerged content", we tested the use of the "Sliding Text Window" as the classification unit.

The motivation for such an approach was found in the n-gram tokenization article (McNamee and Mayfield, 2004), allowing to match parts of text without

Table 4: The sliding window allows to zoom on the significant content for the Complementarity criterion. A few samples from the page of Figure 3.

No.	Content of the window
3	Content The information on this Site has been included in good faith but is for general informational purposes only. It should not be relied on for any specific purpose and no representation or warranty is given as regards its accuracy or completeness. No information on this Site shall constitute an invitation to invest in the Company, nor should it be used as the basis for any investment decision. *The Content of this Site is not intended as medical advice, nor is it recommended as a*
4	Completeness. No information on this Site shall constitute an invitation to invest in the Company, nor should it be used as the basis for any investment decision. *The Content of this Site is not intended as medical advice, nor is it recommended as a substitute for medical advice. You should always seek the advice of your doctor or health care professional regarding any medical condition or treatment.* Neither the Company, its affiliates, nor their respective directors, officers, employees, agents
5	*Substitute for medical advice. You should always seek the advice of your doctor or health care professional regarding any medical condition or treatment.* Neither the Company, its affiliates, nor their respective directors, officers, employees, agents, or representatives are engaged in rendering medical advice. Alere reserves the right to make any changes and corrections to this Site and its Content as and when we consider it appropriate and without notice. Privacy Policy Alere$^{\text{TM}}$Privacy

losing track of context. The size of the window was established empirically. We chose to create a window consisting of 500-characters maximum (limited to the word boundaries) and slid it progressively 250 characters at a time.

Table 4 gives examples of the different passages for the webpage displayed in Figure 3 and highlighted by the blue rectangle. In order for the page to be marked as respecting the criteria, the system needs to detect its presence on at least one window created in such a way for this page. Information related to the *Complementarity* criterion spreads in the rows 3, 4 and 5. This information is in *italics*.

To test the effectiveness of the above-described methods for both the *Date Attribution* and *Complementarity* criteria, we used the same set of 27 websites (+10,000 webpages) as the ones used for the previous experiments (Boyer and Dolamic, 2015). The main motivation for using the same collection and setup (i.e. Nave Bayes classifier in combination with word tokenization) was to be able to compare the system before/after directly while integrating the new approaches customized for the 2 criteria. The convenience sample of 27 health websites was selected to broadly cover HONcode potential and actual sites as follows:

- 9 new, potentially certifiable websites. HONcode experts estimated that these websites do conform to HONcode, but are not yet certified;
- 9 likely non-certifiable websites. The HONcode experts estimated that these websites would not conform to HONcode principles when fully analysed;
- 4 newly certified websites. These websites were recently certified for the first time;
- 5 previously certified HONcode sites chosen because they were awaiting annual reassessment.

In order to perform different research experiments conducted over a few years, we decided to locally retrieve the websites using the HON crawler. The crawling was conducted in April 2014. It should be noted that it was a good approach as several websites do not exist anymore (10% of websites selected closed down).

We have chosen to present the obtained results using various measurements. Apart from giving the standard classification measurements: precision (P), recall (R) and accuracy (A); we added the contingency table values: False Negative (FN), False Positive (FP), True Negative (TN) and True Positives (TP).

4 RESULTS

Table 5 gives the results for the detection of the *Date Attribution* criterion. Manual review has resulted in it being detected for 21 out of 27 websites in the test set (column Manual +). In the case that neither the automated system nor manual review found evidence supporting this criterion, it was considered as true negative (TN), while detection by both manual and automated system is considered a true positive (TP). Websites for which the criterion was detected in the manual review but not by the automated system are considered to be a false negative (FN), while the ones detected by the automated system, but not in the manual review, represent a false positive (FP). In the experiments which yielded the results are presented in this table we compared the results obtained by a machine

Table 5: Results for the Date Attribution detection with different techniques (N=27); Legend: Doc. Document; not compliant; + compliant; TP True Positive; TN True Negative; FP False Positive; FN False Negative; Precision P; Recall R; Accuracy A.

Det. unit/ Met.	Manual eval. 27 web-sites		Date (Attribution) Automated detection						
	+	-	TP	TN	FP	FN	P	R	A
Doc.	21	6	5	6	0	16	100	24	41
Sent.	21	6	20	0	6	1	77	95	74
NER	21	6	19	1	5	2	79	90	74

Table 6: Zoom on supposed incorrect (FP) detection of the attribution criterion with the NER approach.

No.	Date criterion detected by NER method
1	25 August 2003‡
2	Nutrition April 30, 2013
3	last edited Jan 18, 2013‡
4	July 1, 2013‡
5	Thursday, February 21, 2013 from 9a.m.

learning approach (using two different classification units, namely Document (Doc.) and Sentence (Sent.)) to the performance of the Named Entity Recognition (NER).

From the results presented in Table 5, it can be noticed that both Sentence machine learning and the NER approach result in lower precision (77% and 79%) when compared to results obtained by the the machine learning Document approach. However, they also result in significantly higher recall (95% and 90% for Sentence and NER respectively vs. 24% for Doc and accuracy (74% for NER or sentence vs. 41% for Doc.).

Both Sentence and NER approaches resulted in some false positive results (e.g. 6 and 5 respectively). Accounting for the nature of the NER detection, we were able to find the exact values concerning these false positives (FP) detected by the system. These values are given in the Table 6. Manual re-inspection of the content of these documents showed that the values marked by ‡ in this table represent True Positive detection and should have not been classified as FP.

Table 7 provides the results of the comparison between the manual and automated approaches for the *Complementarity* criterion. The legend of the columns in the Table 7 is the same as those in Table 5. For this criterion, we kept the machine learning approach and varied the classification unit in order to deal with "submerged content" without creating too much noise.

Table 7: Results for the Complementarity criterion using different classification units (N=27).

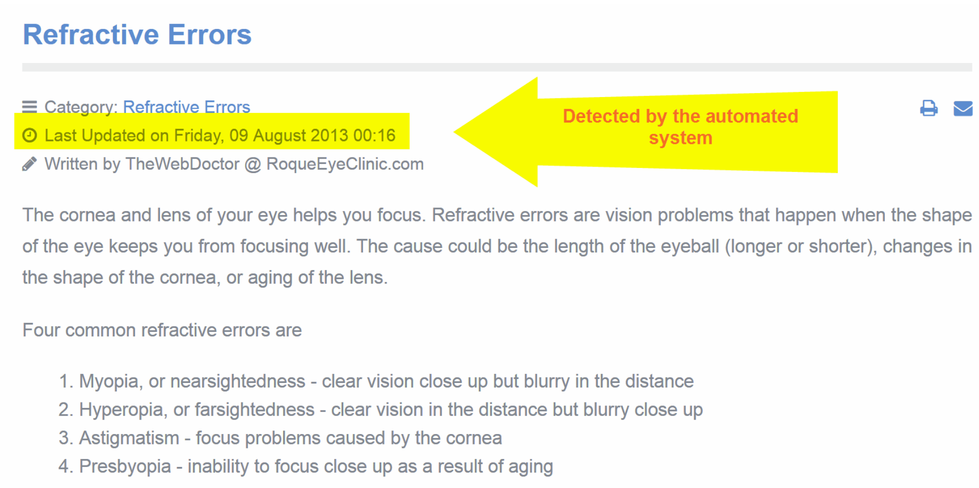
Det. unit	Manual eval. 27 web-sites		Complementarity Automated detection						
	+	-	TP	TN	FP	FN	P	R	A
Doc.	26	1	5	6	0	16	100	19	22
Sent.	26	1	22	0	1	4	95	85	74
Win.	26	1	24	0	1	2	96	92	74

Thus, we used Document (Doc.), Sentence (Sent.) and Sliding Window (Win.) as a classification unit for this purpose. It can be observed in the results presented here that both Sentence and Window approach result in slightly lower precision (96% compared to 100% for Document). On the other hand, the increase in recall for both these approaches, when compared to that of Document is highly significant (85% and 92% for Sentence and Window respectively vs. 19% for Doc.). This is also the case for the accuracy.

5 DISCUSSION

Even though with the machine learning approach to *Date Attribution* detection, when a sentence is used as classification unit, the results in the performance are comparable to that of the NER method, closer inspection showed two main problems. Similarly to that for the Privacy criterion described in (Boyer and Dolamic, 2015), this approach creates a lot of noise for the *Date Attribution* criterion as well. For example, on the webpage related to advertisement policy, http://www.webmd.com/about-webmd-policies/about-advertising-policy, when using Sentence as a classification unit, the system detects the Privacy criteria only because of the sentence: *"WebMD may change this policy at any time in its sole discretion by posting a revised policy to the applicable WebMD Property"*. This is a consequence of the very high probability of the term "policy" for the Confidentiality criterion. Similarly, *Date Attribution* is detected in the sentence *"Nosebleeds that last more than 30 minutes require medical attention"* based on the term "last". It is important to state that neither of these principles is detected in the windows containing these sentences.

The second problem remains the system's inability to detect the digit only dates. Both machine learning and the NER approach detect the date on the page http://roqueeyeclinic.com/roque-eye-clinic-patient-information/eye-conditions/refractive-errors illustrated in Figure 4. Unlike the machine learning approach, the NER is also capable of detecting the only digit date format 2014-07-



Refractive Errors

≡ Category: Refractive Errors

Last Updated on Friday, 09 August 2013 00:16

Written by TheWebDoctor @ RoqueEyeClinic.com

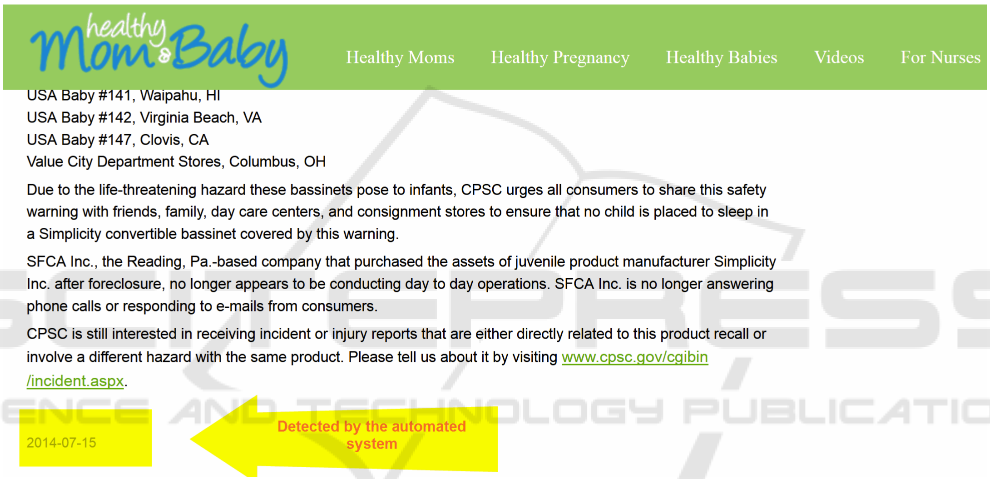
Detected by the automated system

The cornea and lens of your eye helps you focus. Refractive errors are vision problems that happen when the shape of the eye keeps you from focusing well. The cause could be the length of the eyeball (longer or shorter), changes in the shape of the cornea, or aging of the lens.

Four common refractive errors are

1. Myopia, or nearsightedness - clear vision close up but blurry in the distance
2. Hyperopia, or farsightedness - clear vision in the distance but blurry close up
3. Astigmatism - focus problems caused by the cornea
4. Presbyopia - inability to focus close up as a result of aging

Figure 4: Date Attribution criteria detected by both NER and machine learning approaches.



USA Baby #141, Waipahu, HI
USA Baby #142, Virginia Beach, VA
USA Baby #147, Clovis, CA
Value City Department Stores, Columbus, OH

Due to the life-threatening hazard these bassinets pose to infants, CPSC urges all consumers to share this safety warning with friends, family, day care centers, and consignment stores to ensure that no child is placed to sleep in a Simplicity convertible bassinet covered by this warning.

SFCA Inc., the Reading, Pa.-based company that purchased the assets of juvenile product manufacturer Simplicity Inc. after foreclosure, no longer appears to be conducting day to day operations. SFCA Inc. is no longer answering phone calls or responding to e-mails from consumers.

CPSC is still interested in receiving incident or injury reports that are either directly related to this product recall or involve a different hazard with the same product. Please tell us about it by visiting www.cpsc.gov/cgibin/incident.aspx.

2014-07-15

Detected by the automated system

Figure 5: Date attribution criterion detected only when using the NER method.

15 without any textual clarification in the page http://www.health4mom.org/reannouncement-of-simplicity-bassinets-recall illustrated in Figure 5.

For the *Complementarity* criterion, the results obtained show that the Sentence and Window approaches yield comparable outcomes. However, taking a closer look at the result details shows that the Sentence approach results in a higher number of detections based on a small number of terms. Thus, this criterion is detected in the sentence *"Point of care testing is often described as the preferred method of INR testing because it allows the healthcare professional to read and interpret the results while the patient waits which saves time and facilitates effective counselling"* due to the terms "healthcare" and "professionals". Using the window approach also solves this problem.

Apart from the two HONcode criteria examined in detail in this article, *Date Attribution* and *Confidentiality*, we tested the effect of the Sliding Window classification unit on the other 6 criteria. Our results prove that this method is also efficient for criteria such as *Advertising policy* or *Financial disclosure* providing the following improvement.

6 CONCLUSION

Based on the results presented in this article, we can conclude that in order to achieve the optimal results for all HONcode criteria, it is necessary to perform customised criteria-optimisation of the system. The results show that using the Sliding Window as the classification unit instead of Document proves to be a good choice not only for the *Complementarity* criterion but also for the *Advertising policy* or textitFinancial disclosure ones. The common characteristic of these criteria is that they are, for many websites, present on the same webpage as in the following example http://ptinr.com/terms-and-conditions.

On the other hand, for the *Date Attribution* criterion, changing the classification somewhat improved the system. However, for the problem of the date in digit only format, the machine learning approach proved to be ineffective. Consistent with results reported in (Vishnyakova et al., 2014) relating to the determination of duration in clinical trials, a simple Named Entity Recognition tool, trained on the HONcode specific data, proved to be the right solution to handle numeric sequences.

ACKNOWLEDGEMENTS

The research is conducted in the scope of the European project Kconnect and funded by this project (2015-2018, project No. 644753) in the continuation of the European project KHRESMOI (2010-2014, project No. 257528).

REFERENCES

Aphinyanaphongs, Y., Tsamardinos, I., Statnikov, A., Hardin, D., and Aliferis, C. (2005). Text categorization models for high-quality article retrieval in internal medicine. *J Am Med Inform Assoc*, 12(2):207–216.

Boyer, C., Baujard, V., Nater, T., Scherrer, J., and Appel, R. (1999). The health on the net code of conduct for medical and health-related web sites: three years on. *J Med Internet Res*, 1(Suppl 1):e99:e99.

Boyer, C. and Dolamic, L. (2014). Feasibility of automated detection of honcode conformity for health-related websites. *IJACSA*, 5(3):69–74.

Boyer, C. and Dolamic, L. (2015). Automated detection of honcode website conformity compared manual detection: An evaluation. *J Med Internet Res*, 17(6):e135.

Boyer, C., Hajic, J., Hanbury, A., Kirtz, M., Pletneva, N., Schneller, P., Stefanov, V., and Uresova, Z. (2014). D10.3: Report on the extensive tests with the final search system. Khresmoi public deliverable. Accessed on : 25.08.2015. *http://khresmoi.eu/assets/Deliverables/WP10/KhresmoiD103.pdf*.

Griffiths, K., Tang, T., Hawking, D., and Christensen, H. (2005). Automated assessment of the quality of depression websites. *J Med Internet Res*, 7(5):e59.

Humphrey, T. (2009). Internet users now spending an average of 13 hours a week online. Accessed on: 08.01.2012. http://news.harrisinteractive.com/profiles/investor/ResLibraryView.asp?BzID=1963& ResLibraryID=35164 &Category=1777.

Ilic, D., Bessel, T., Silagy, C., and Green, S. (2003). Specialized medical search engines are no better than general search-engines in sourcing consumer information about androgen deficiency. *Hum Reprod.*, 18(3):557–561.

McNamee, P. and Mayfield, J. (2004). Character n-gram tokenization for european language text retrieval. *Information Retrieval*, 7(1-2):7397.

OpenNLP (2015). Apache opennlp developer documentation. Accessed on : 25.08.2015. http://opennlp.apache.org/documentation/manual/opennlp.html#tools.namefind.recognition.

van Straten, A., Cuijpers, P., and Smits, N. (2008). Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: Randomized controlled trial. *J Med Internet Res*, 10(1):e7.

Vishnyakova, D., Gobeill, J., Oezdemir-Zaech, F., Kreim, O., Vachon, T., Cladé, T., Haenning, X., Mikhailov, D., and Ruch, P. (2014). Electronic processing of informed consents in a global pharmaceutical company environment. *MIE*, pages 995–999.

Williams, K. and Calvo, R. (2002). A framework for document categorization. In *Proceedings of the Seventh Australasian Document Computing Symposium.*