Using Factual Density to Measure Informativeness of Web Documents

Christopher Horn[*], *Alisa Zhila*[†], *Alexander Gelbukh*[†], *Roman Kern*[*], *Elisabeth Lex*[*]

[*]Know-Center GmbH, Graz, Austria
{chorn,elex,rkern}@know-center.at

[†]Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City, Mexico
alisa.zhila@gmail.com, gelbukh@gelbukh.com

ABSTRACT

The information obtained from the Web is increasingly important for decision making and for our everyday tasks. Due to the growth of uncertified sources, blogosphere, comments in the social media and automatically generated texts, the need to measure the quality of text information found on the Internet is becoming of crucial importance. It has been suggested that factual density can be used to measure the informativeness of text documents. However, this was only shown on very specific texts such as Wikipedia articles. In this work we move to the sphere of the arbitrary Internet texts and show that factual density is applicable to measure the informativeness of textual contents of arbitrary Web documents. For this, we compiled a human-annotated reference corpus to be used as ground truth data to measure the adequacy of automatic prediction of informativeness of documents. Our corpus consists of 50 documents randomly selected from the Web, which were ranked by 13 human annotators using the MaxDiff technique. Then we ranked the same documents automatically using ExtrHech, an open information extraction system. The two rankings correlate, with Spearman's coefficient $\rho = 0.41$ at significance level of 99.64%.

KEYWORDS: quality of texts, Web, fact extraction, open information extraction, informativeness, natural language processing.

1 Introduction

Assessment of information quality becomes increasingly important because nowadays decision making is based on information from various sources that are sometimes unknown or of questionable reliability. Besides, a large part of the information found in the Internet has low quality: the Internet is flooded with meaningless blog comments, computer-generated spam, and documents created by copy-and-paste that convey no useful information.

As one might assume, talking about the quality of the Internet content on the whole is too general and practically impossible, because the content on the Web is of an extremely versatile form and serves to very different needs. It is unreasonable to compare the quality of an online encyclopedia to the quality of a photo storing resource because the assessment of the quality of any object depends on the purpose to which the object serves. Hence, we restrict ourselves to the scope of text documents and assume that the general purpose of a text document is to inform a reader about something. Therefore, the quality of a text document can be related to its *informativeness*, i.e. the amount of useful information contained in a document. (Lex et al., 2012) suggest that informativeness of a document can be measured through factual density of a document, i.e. the number of facts contained in a document, normalized by its length.

Due to the lack of a standard corpus, previous works on the estimation of Web quality concidered only Wikipedia articles, in fact assessing their informativeness. No special studies were performed about human judgement on text informativeness. Therefore, (Lex et al., 2012; Blumenstock, 2008; Lex et al., 2010; Lipka and Stein, 2010) considered Wikipedia editors' choice of *featured* and *good* articles as a reasonable extrapolation of high judgement on their informativeness. (Lex et al., 2012) showed the feasibility of factual density application as measurement of informativeness on the base of automatic prediction of the featured/good Wikipedia articles.

In this work we have conducted experiments to estimate the adequacy of application of factual density to informativeness evaluation in the "real" Internet, i.e. not limited to particular web-sites with a particular form of content but rather covering a wide variety of web-sources, which a user could browse through while looking for information. For this purpose we created a dataset of 50 randomly selected documents in Spanish language from CommonCrawl corpus (Kirkpatrick, 2011), which is a large extraction of texts from the Internet. We assessed factual density automatically using an open information extraction system for Spanish language, ExtrHech (Zhila and Gelbukh, 2013), which is adequate for Web-scale applications. Further, 13 human annotators ranked 50 documents according to their informativeness using the MaxDiff (Louviere and Woodworth, 1991) technique. The automatic ranking produced by ExtrHech system correlates with the ground truth ranking by human annotators with Spearman's ρ coefficient of 0.41 (coinciding rankings would have 1, and the random baseline is 0.018).

The paper is organized as follows. In section 2 we review the related work on the Web text quality evaluation as well as the open information extraction as a method for factual density estimation. Section 3 describes the dataset created in this work and the human annotation procedure. The method for automatic factual density estimation and a brief description of the architecture of ExtrHech system is given in section 4. The experiment and its results are presented in section 5. In section 6 we give a brief discussion of a possible expansion of our text quality measuring method to other languages. Section 7 concludes this paper with an

overview of the presented work and proposals for future work.

2 Related work

Evaluation of the quality of Web text content has been mainly performed with metrics capturing content quality aspects like objectivity (Lex et al., 2010), content maturity, and readability (Weber et al., 2009). These methods are based on selection of appropriate features for document presentation. For example, in (Lex et al., 2010) stylometric features were used to assess the content quality. Character trigram distributions were exploited in (Lipka and Stein, 2010) to identify high quality *featured/good* articles in Wikipedia. (Blumenstock, 2008) considered simple word count as an indicator for the quality of Wikipedia articles. (Lex et al., 2012) proposed factual density as a measure of document informativeness and showed that it gives better results for Wikipedia articles than other methods. Wikipedia articles were taken into consideration mainly due to the lack of a standard corpus in this field of work. For evaluation purposes, those Wikipedia articles that have the featured article or good article template in the wikitext were considered to be of a high quality or more informative. No specially designed human annotation or evaluation was involved, and no scale or ranking of informativeness was introduced.

To asses factual density of a text document, (Lex et al., 2012) apply the open information extraction (Open IE) methods. Open IE is the task of extracting relational tuples representing facts from text without requiring a pre-specified vocabulary or manually tagged training corpora. A relational tuple in most current Open IE systems is a triplet consisting of two arguments in the form of noun phrases and a relational phrase expressed as a verb phrase. Consequently, a fact is presented as a triplet like $f = (Mozart, was\ born\ in, Salzburg)$. The abscence of a pre-specified vocabulary or a list of predetermined relations makes such systems scalable to a broader set of relations and to a far larger corpora as the Web (Etzioni et al., 2008).

Nevertheless, the output of the Open IE systems contains a large portion of uninformative and incoherent extractions that can lead to overestimating of the factual density in a document. (Fader et al., 2011) show that a part-of-speech based system like ReVerb with simple syntactic and lexical constraints increases the precision of the output and needs no annotating or training effort for its implementation compared to other Open IE systems like TextRunner (Banko et al., 2007), WOE^{pos}, and WOE^{parse} (Wu and Weld, 2010), which all include a relation learning step.

3 Building the Ground Truth Dataset

Since no special corpus for informativeness evaluation previously existed, we aimed at creation of such a corpus of texts extracted from the Internet broader than Wikipedia.

3.1 The dataset

For the purpose of evaluation of the factual density as a measure of informativeness, we needed to create a dataset that would be a reasonable projection of texts on the Web and small enough to be able to conduct the experiment with available resources. To create our dataset we performed the following steps:

- We used a 1 billion page subset from the CommonCrawl corpus (Kirkpatrick, 2011) from 2012, which is a corpus of the web crawl data composed of over 5 billion web

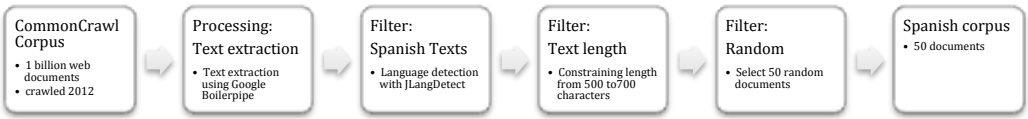

Figure 1: Process of corpus generation

pages, as an initial source of Web texts. From that corpus, we extracted textual content of websites using Google's Boilerpipe framework (Kohlschütter et al., 2010).

- For each article, the language was detected using JLangDetect (Nakatani, 2011).
- From this dataset, we randomly selected 50 documents in Spanish. In order to avoid the length-based bias on the human annotation stage described in the next subsection (e.g. users might tend to rate longer texts as more informative than shorter ones), we constrained the text length to range from 500 to 700 characters.

In the end, we formed a corpus of 50 text documents in Spanish of similar length that represent a random sample of the textual content from the Web. Figure 1 shows the process.

We would like to emphasize that not all textual content presented on the Web is coherent text. The texts encountered on the Internet can consist of pure sequences of keywords, or be elements of web-page menus, for example, " *For more options click here Leave your comment CAPTCHA*". Lists and instructions are another common form of texts, characterized by incomplete sentences normally starting with a verb in the infinitive or imperative form, e.g. "*To open a file: – click the Microsoft Office button; – select Open*". Texts can be sets of short comments or tweets that also tend to be incomplete sentences often lacking grammatical correctness. Commercials and announcements also typically consist of incomplete sentences, e.g. "*Information about the courses, dates, and prices*", numbers, e.g. "*$65 $75 $95 All prices in US dollars*", and telephone numbers and addresses. We manually performed a rough classification of the texts from our dataset shown in Table 1. Since we used only short documents for the experiment, each document mainly corresponded to only one text type.

In the current work we did not do any additional pre-processing for text type detection. This was not done for several reasons. First, we want to keep the system as simple and fast as possible for the purpose of scalability to large amounts of text. Next, we believe that the factual density approach presented in the paper will be appropriate for automatic detection of incoherent and uninformative texts. Consequently, there will be no need for additional filtering.

3.2 Ground Truth Ranking by Human Annotators

To overcome the lack of a standard corpus in the field of web text informativeness assessment, we formed a ground truth ranking of the documents based on inquiry of 13 human annotators. All human annotators are natively Spanish speaking people with graduate level of education. For the questionnaire, we opted for the MaxDiff (Maximum Difference Scaling) technique (Louviere and Woodworth, 1991). According to MaxDiff technique, instead of ranking all items at once, a participant is asked to choose the best and the worst item from a subset of items at a time. This procedure is repeated until all items are covered.

Type of text	# of docs	Characteristics
keywords	2	sequence of nouns with no verbs
web page menu	1	short phrases, verbs
commercials, announcements	18	addresses, phone numbers, prices, imperatives
coherent narrative: descriptions, news	13	full sentences with subjects, verbs, and objects
comments, tweets	6	short sentences, lacking grammatical correctness
instructions, lists	9	phrases starting with infinitives or no verbs
incorrectly detected language	1	impossible to POS-tag for the system and to read for human annotators

Table 1: Classification of the documents in the dataset by the types of text content

MaxDiff is considered to be a strong alternative to standard rating scales (Almquist and Lee, 2009). The advantage of the MaxDiff technique is that it is easier for a human to select the extremes of a scale rather than to produce a whole range of scaling. Consequently, MaxDiff avoids the problems with scale biases and is more efficient for data gathering than the simple pairwise comparison.

For this purpose, we created a set S of MaxDiff questions. Each question, which will subsequently be called Q, contained 4 different documents d. Four items is few enough for a participant to be able to concentrate and to make a decision, and large enough to be able to cover all 50 documents in a reasonable number of questions.

The set S was created as follows:

- First, we calculated all possible combinations of how 4 documents can be chosen from 50 documents. This resulted in 230,300 possible combinations.
- Then, in a loop, we picked up one random question Q from the set of combinations and added it to the resulting set S, until every document d was included three times in the set S. This ensures that every document is compared at least three times with other documents. Once finished, the resulting set S contained 103 questions Q, which we used for the MaxDiff inquiry.

Further, we created a MaxDiff questionnaire system that displayed one question Q' (= 4 documents d') at a time to a participant. A participant was asked to select the most informative document and the least informative document from the set of 4 documents. The interface of the questionnaire system is shown in Figure 2. In the experiment, the instructions and the documents were given to the annotators in Spanish language. In Figure 2 they are translated into English for convenience of the reader of this article. The selection of the most and the least informative documents was based on the intuitive understanding of the notion of "informativeness" by the participants. Each of the participants rated 25 questions on average. The system ensured that each question Q is (i) rated three times in total and

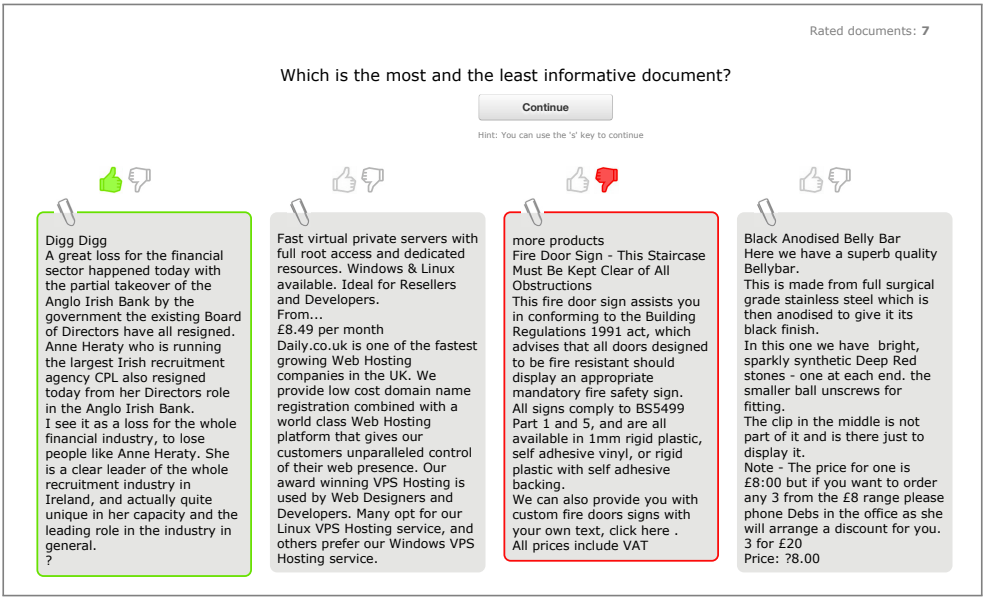

Figure 2: Screenshot of the MaxDiff questionnaire tool

(ii) rated by different users. This resulted in 309 ratings, with each question being answered 3 times.

We applied the MaxDiff technique to the answeres obtained from the user-study. The rank for each document d was calculated proportionally to its MaxDiff scoring $\text{Score}_{\text{maxdiff}}(d)$, which was calculated using the following formula:

$$\text{Score}_{\text{maxdiff}}(d) = R_{\text{pos}}(d) - R_{\text{neg}}(d), \tag{1}$$

where $R_{\text{pos}}(d)$ is the number of positive answers and $R_{\text{neg}}(d)$ is the number of negative answers for the same document d.

After calculating the score for each document d, we formed a ground truth ranking of the 50 documents in our dataset.

4 Automatic Ranking

The aim of this work is to study the feasibility of automatic factual density estimation for informativeness measurement for text documents on the Web. This section describes the procedure for the automatic ranking.

4.1 Factual Density

In the factual density approach to web informativeness assessment, each text document is charecterized by the factual density feature. To calculate the value of factual density, first, the simple fact count is determined for each document d using the Open IE method. That means that only direct information on the number of facts, i.e. fact count $fc(d)$, obtained from a text resource d is taken into account. It is obvious that the fact count in a document

is correlated with its length, i.e. longer documents would tend to have more facts than shorter ones. To overcome this dependency, factual density $fd(d)$ is calculated as a fact count in a document $fc(d)$ divided by the document size $size(d)$: $fd(d) = \frac{fc(d)}{size(d)}$.

In this work we used the Open IE system for Spanish ExtrHech, which is described in the next subsection, to determine the fact count in a document. The length of a document was calculated as a number of characters including white spaces.

4.2 Fact Extraction

ExtrHech is a POS-tag based Open IE system for Spanish with simple syntactic and lexical constraints (Zhila and Gelbukh, 2013). The system takes a POS-tagged text as an input. For POS-tagging we used Freeling-2.2 (Padró et al., 2010). Then, it imposes syntactic constraints in the form of regular expressions to each sentence. Since in our framework a fact is a triplet of two arguments and a relation between them, the system requires a relation to be a verb phrase that appears between two arguments expressed as noun phrases. Appropriate syntactic constrains detect adequate verb phrases, resolve coordinating conjunctions for relations and noun phrase arguments, correctly treat participle and relative clauses. In the current version of ExtrHech system, lexical constraints limit the length of relational phrases to prevent overspecifying of a relation. Specifically for Spanish language, the current version is adjusted to EAGLES POS-tag set and properly treats reflexive pronouns for verb phrases.

At the current stage, ExtrHech system has various limitations. It does not resolve anaphora, zero subject construction, and free word order that occur in Spanish. Yet despite of these limitations, its precision and recall is comparable to that of ReVerb system for English language, which was used in previous work on the text quality assessment for Wikipedia articles (Lex et al., 2012).

It is also shown in (Fader et al., 2011) that, although deeper syntactic analysis would increase the precision of syntactic and lexical constraint based Open IE systems, it would inevitably slow down execution time that highly affects the overall time of processing for a Web-scale corpus.

5 Experiment and Results

In this work we conducted an experiment to study the appropriateness of factual density measure for assessment of web text informativeness. In order to prove the hypothesis, we compared the ranking based on automatic factual density scoring to the ground truth ranking based on the MaxDiff inquiry of human annotators.

To form the factual density based ranking, 50 documents were fed into a pipeline: Freeling-2.2 POS-tagger, ExtrHech Open IE system, and a script for factual density calculation shown in Figure 3.

Then, each document d was ranked according to its factual density scoring $\text{Score}_{\text{factdens}}(d)$:

$$\text{Score}_{\text{factdens}}(d) = \text{fc}(d)/\text{size}(d), \quad (2)$$

where $fc(d)$ is the fact count for a document d, and $size(d)$ is its length in characters including white spaces.

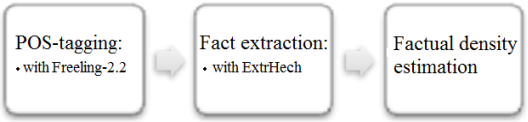

Figure 3: Diagram of the factual density estimation

Human annotator ranking was formed as described in Section 3.2. The rankings are shown in Table 2, where HA rank is the human annotator ranking and FD rank is the factual density ranking.

Doc ID	HA rank	FD rank	Doc ID	HA rank	FD rank
1	1	3	26	24.5	40
2	2.5	11	27	27	14
3	2.5	29.5	28	29.5	15
4	4.5	2	29	29.5	22
5	4.5	7	30	29.5	32
6	6	6	31	29.5	16
7	7	12	32	33	24
8	8.5	33	33	33	9
9	8.5	1	34	33	43
10	10	5	35	37	47
11	11.5	27	36	37	47
12	11.5	8	37	37	35
13	14.5	42	38	37	34
14	14.5	39	39	37	10
15	14.5	19.5	40	40	17
16	14.5	41	41	41.5	47
17	19	21	42	41.5	47
18	19	4	43	44	18
19	19	29.5	44	44	47
20	19	36	45	44	37
21	19	38	46	46	26
22	22	23	47	47	19.5
23	24.5	13	48	48	25
24	24.5	47	49	49	31
25	24.5	47	50	50	28

Table 2: Human annotator ranking and factual density ranking

Once the rankings were scored, we applied various statistical measures to calculate the correlation between them. Table 3 shows the results of the statistical evaluation using Spearman's ρ coefficient, Pearson product-moment correlation r, and Kendall's rank correlation τ with the corresponding levels of significance. All these correlation coefficients may vary between 1 for coinciding rankings and -1 for completely opposite rankings. Random baseline for Spearman's ρ is 0.018. The coefficients should be significantly positive for the rankings to be considered correlated.

In our work we obtained Spearman's ρ as high as 0.41, Pearson's r of 0.38, and Kendall's τ of 0.29. Since all measures give significantly positive correlations with the significance level equal to or higher than 99.49%, we can conclude that medium correlation significantly exists between the two rankings. Consequently, the obtained result show that the factual density measure proves to be feasible as a measure of informativeness for text content on the Web.

Method	Value	P-Value	Significance Level
Spearman's ρ	0.404	0.00365	99.636%
Pearson's r	0.390	0.00514	99.486%
Kendall's τ	0.293	0.00347	99.653%

Table 3: Result of correlation tests between factual density algorithm ranking and human annotator ranking for 50 document dataset

6 Discussion: Applicability to Other Languages

In the current work we studied the adequacy of the factual density estimation as a quality measure for arbitrary Web texts on the material of Spanish language. As mentioned in Section 2, the text quality measuring through factual density estimation with an Open IE system for English was shown to be adequate for English language Wikipedia articles (Lex et al., 2012). Our future goal is to show the adequacy of the method for arbitrary English texts as well, especially social media texts.

Moreover, the approach outlined in Sections 3 and 4 can be applied to texts in any languages, to which the fact extraction paradigm described in Section 4.2 is applicable. To the best of our knowledge, so far such Open IE methods have been elaborated and corresponding fact extraction systems have been implemented only for English (Fader et al., 2011) and Spanish (Zhila and Gelbukh, 2013). Generally, this approach to fact extraction requires a POS-tagger and relies on the assumption of a single prevailing word order. Therefore, this fact extraction technique can be similarly implemented for languages with a fixed or dominating word order and for which reliable POS-taggers exist. This is the case for most European languages, e.g. $97-98\%$ POS-tagging accuracy for the languages supported by Freeling-3.0 package (Padró and Stanilovsky, 2012). However, currently POS-tagging in other languages performs with lower accuracy, e.g. $\sim$88% for Bengali (Dandapat et al., 2007) or 94.33% for Chinese (Ma et al., 2012). The issue of how POS-tagging accuracy affects performance of a fact extraction system and, consequently, what effect it has on the text quality measuring is out of the scope of the paper and is an interesting direction for future work.

Next, the sensitivity of the fact extraction method to the word order questions the feasibility of text quality measuring using similar approach for languages with flexible word order, e.g. Chinese. We believe that further research in this area will answer these questions.

7 Conclusions and Future Work

In this work we moved the research on informativeness of the Web from Wikipedia articles to the arbitrary Web texts. We formed a dataset of 50 Spanish documents extracted from the Web. Then, we presented a MaxDiff questionnaire to 13 human annotators to be able to form a ground truth ranking of informativeness of Web text documents based on the

human judgements. Further, with the use of the open information extraction system ExtrHech, we calculated the factual density of the documents and estimated the corresponding ranking. In the end, we showed that significant positive correlation exists between the automatically generated ranking based on factual density and the ranking based on human judgements (Spearman's rank correlation of 0.404 at a significance level of 99.636%, n=50). Consequently, we conclude that the factual density measure can be applied for adequate assessment of the informativeness of textual content on the Web.

In future work, we plan to conduct similar experiments for other languages to analyze how language affects factual density estimation. Next, we are going to analyze how a text form of a document, i.e. lists, instructions, announcements, comments, news, etc., is related to its informativeness. We also want to consider how a topic of a text document and human annotator's personal interests might change his or her judgements about the informativeness of a document. For example, whether a person interested in sports would always choose sport-related articles as more informative. Another branch of our research is dedicated to the improvement of the fact extraction method and increase of the precision of the factual density estimation.

Acknowledgements

This work has been funded by the European Commission as part of the WIQ-EI project (project no. 269180) within the FP7 People Programme and Instituto Politécnico Nacional grant SIP 20131702 .

References

Almquist, E. and Lee, J. (2009). What do customers really want? `http://hbr.org/2009/04/what-do-customers-really-want/ar/1`. [last visited on 09/04/2013].

Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., and Etzioni, O. (2007). Open information extraction from the web. In *IN IJCAI*, pages 2670–2676.

Blumenstock, J. E. (2008). Size matters: word count as a measure of quality on wikipedia. In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 1095–1096, New York, NY, USA. ACM.

Dandapat, S., Sarkar, S., and Basu, A. (2007). Automatic part-of-speech tagging for bengali: an approach for morphologically rich languages in a poor resource scenario. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 221–224, Stroudsburg, PA, USA. Association for Computational Linguistics.

Etzioni, O., Banko, M., Soderland, S., and Weld, D. S. (2008). Open information extraction from the web. *Commun. ACM*, 51(12):68–74.

Fader, A., Soderland, S., and Etzioni, O. (2011). Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1535–1545, Stroudsburg, PA, USA. Association for Computational Linguistics.

Kirkpatrick, M. (2011). New 5 billion page web index with page rank now available for free from common crawl foundation. `http://readwrite.com/2011/11/07/common_crawl_foundation_announces_5_billion_page_w`. [last visited on 25/01/2013].

Kohlschütter, C., Fankhauser, P., and Nejdl, W. (2010). Boilerplate detection using shallow text features. In *Proceedings of the third ACM international conference on Web search and data mining*, WSDM '10, pages 441–450, New York, NY, USA. ACM.

Lex, E., Juffinger, A., and Granitzer, M. (2010). Objectivity classification in online media. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, HT '10, pages 293–294, New York, NY, USA. ACM.

Lex, E., Voelske, M., Errecalde, M., Ferretti, E., Cagnina, L., Horn, C., Stein, B., and Granitzer, M. (2012). Measuring the quality of web content using factual information. In *Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality*, WebQuality '12, pages 7–10, New York, NY, USA. ACM.

Lipka, N. and Stein, B. (2010). Identifying featured articles in wikipedia: writing style matters. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 1147–1148, New York, NY, USA. ACM.

Louviere, J. J. and Woodworth, G. (1991). Best-worst scaling: A model for the largest difference judgments. Technical report, University of Alberta.

Ma, J., Xiao, T., Zhu, J. B., and Ren, F. L. (2012). Easy-first chinese pos tagging and dependency parsing. In *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers*, pages 1731–1746. Indian Institute of Technology Bombay.

Nakatani, S. (2011). Language detection library for java. http://code.google.com/p/language-detection/. [last visited on 25/01/2013].

Padró, L., Reese, S., Agirre, E., and Soroa, A. (2010). Semantic services in freeling 2.1: Wordnet and ukb. In Bhattacharyya, P., Fellbaum, C., and Vossen, P., editors, *Principles, Construction, and Application of Multilingual Wordnets*, pages 99–105, Mumbai, India. Global Wordnet Conference 2010, Narosa Publishing House.

Padró, L. and Stanilovsky, E. (2012). Freeling 3.0: Towards wider multilinguality. In Chair), N. C. C., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).

Weber, N., Schoefegger, K., Bimrose, J., Ley, T., Lindstaedt, S., Brown, A., and Barnes, S.-A. (2009). Knowledge maturing in the semantic mediawiki: A design study in career guidance. In *Proceedings of the 4th European Conference on Technology Enhanced Learning: Learning in the Synergy of Multiple Disciplines*, EC-TEL '09, pages 700–705, Berlin, Heidelberg. Springer-Verlag.

Wu, F. and Weld, D. S. (2010). Open information extraction using wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 118–127, Stroudsburg, PA, USA. Association for Computational Linguistics.

Zhila, A. and Gelbukh, A. (2013). Comparison of open information extraction for english and spanish. In *Accepted for Proceedings of the 19th International Computational Linguistics Conference Dialogue*, Dialogue 2013.