Information Nutrition Labels: A Plugin for Online News Evaluation

Kevin Vincentius   **Birte Högden**   **Claudia Schwenger**
Piyush Aggarwal   **Neelu Madan**   **Farid Muradov**
Ali Şahan   **Anusha Bangaru**   **Ahmet Aker**

University of Duisburg-Essen
`a.aker@is.inf.uni-due.de`

Abstract

In this paper we present a browser plugin *NewsScan* that assists online news readers in evaluating the quality of online content they read by providing *information nutrition labels* for online news articles. In analogy to groceries, where nutrition labels help consumers make choices that they consider best for themselves, information nutrition labels tag online news articles with data that help readers judge the articles they engage with. This paper discusses the choice of the labels, their implementation and visualization.

1 Introduction

Nowadays, the amount of online news content is immense and its sources are very diverse. For the readers and other consumers of online news who value balanced, diverse and reliable information, it is necessary to have access to methods of evaluating the news articles available to them.

This is somewhat similar to food consumption where consumers are presented with a huge variety of alternatives and therefore face the challenge of deciding what is good for their health. This is why food packages come with nutrition labels that guide the consumers in their decision making. Taking this analogy, Fuhr and colleagues (2018) discuss the idea of implementing information nutrition labels for news articles. They propose to label every online news article with information nutrition labels that describe the ingredients of the article and give readers a chance to make an informed judgment about what they are reading. The authors discuss nine different information nutrition labels: factuality, readability, virality, emotion/sentiment, opinion/subjectivity/objectivity, controversy, authority/credibility/trust, technicality and topicality. Gollub and colleagues (2018) categorize these labels into fewer dimensions. Their aim is to establish group labels that are easily understood by readers. However, both studies do not go beyond discussing, proposing and grouping the labels.

In this work we actually implement information nutrition labels and deliver them as a browser plugin that we call *NewsScan*. Therefore we provide a basis for evaluating how well labels describe the online news content and for investigating how useful they are to real users for making decisions about whether to read the news and whether to trust its content. To avoid biasing the user in any way with respect to the consumption of an article, the information is solely presented but not interpreted. Judgments about news sources should be made by users themselves. Whether an article will be read or discarded depends on the user's own weighing of importance of the information nutrition labels.

The plugin supports the reader in these tasks through easy-to-understand visualizations of the labels. In this paper we discuss the methods behind the label computation (Section 2) and the design of the user interface (Section 3).

2 Information nutrition labels

NewsScan implements six information nutrition labels: *source popularity, article popularity, ease of reading, sentiment, objectivity and political bias*. *Ease of reading, sentiment* and *objectivity* have been proposed by Fuhr et al. (2018). We propose to add three more nutrition labels: Source as well as article popularity and political bias. Similar to food nutrition labels the information nutri-

tion labels aim to provide the reader some base to judge about the reliability of the article's content. The credibility nutrition label proposed by Fuhr et al. (2018), for instance, is able to give the reader the indication whether e.g. the source where the article come from is credible or not. However, the credibility label entails already a judgment. It already sums some pieces of information and makes conclusion based on them. We think instead of providing the reader such a judgment the user might be better informed when we provide information that are possible bases for computing e.g. the credibility label. The proposed three new labels aim this purpose, i.e. providing enough details to enable the user to make an informed judgment about an articles content.

In the following we describe the nutrition labels currently implemented within the plugin.

2.1 Source popularity

The label *source popularity* encompasses two dimensions: the reputation of the news source and its influence.

The reputation of a source is analyzed using the **Web of Trust Score**[1]. This score is computed by an Internet browser extension that helps people make informed decision about whether to trust a website or not. It is based on a unique crowdsourcing approach that collects rating and reviews from global community of millions of users who rate and comment on websites based on their personal experiences.

The influence of a source is computed using Alexa Rank, Google PageRank[2] and popularity on Twitter.

Alexa Rank is a virtual ranking system set by Alexa.com (a subsidiary of Amazon) that audits and publishes the frequency of visits on various websites. The Alexa ranking is the geometric mean of reach and page views, averaged over a period of three months.

Google PageRank is a link analysis algorithm that assigns a numerical weight to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of measuring its relative importance within the set.

Twitter Popularity is calculated as an average of the scores for the following two metrics:

- **Followers Count**: This gives the amount of users that are following a source.
- **Listed Count**: This indicates the number of memberships of the source to different topics. It is based on the user's activity to add/remove the source from their customized list. The higher it is, the more diverse the source is.

An overall *source popularity* score shown to the user is calculated by averaging these four metrics. However, when the icon card is flipped the user can also get detailed information about each of the above scores.

2.2 Article popularity

$$Popularity = a\,log(bx + 1) \qquad (1)$$

where *x* is the average amount of tweets per hour, so that the article popularity is *0* when *x* is *0*. The most popular article we found had around 23 tweets per hour in its peak 24 hours. This is used as a reference value, i.e. an article must have this many tweets to reach a score of 100. The logarithmic function is used because the output has to be scaled properly. For example, an article with five tweets per hour is still relatively popular, even though it is just a fraction of the reference score. Choosing a large value for *b* will make the function close to being linear, which will cause even the relatively popular articles to have low scores. A small *b* will make the function more curved. If *b* is too big however, any article with a decent amount of tweets will have a score very close to 100. *b* is chosen empirically to be *1* so that the scores are distributed well between 0 and 100 over a variety of typical news articles. *a* is determined to be *73* to give the reference article a score of 100.

2.3 Ease of reading

As described by Schwarm and Ostendorf (2005) the readability level is used to characterize the educational level a reader needs to understand a text. This topic has been in research since 1930 and several automatic solutions have been proposed to determine the readability level of an input text (Vajjala and Meurers, 2013; Xia et al., 2016; Schwarm and Ostendorf, 2005). The core concept in these studies is to use machine learning along with feature engineering covering lexical, structural, and heuristic based features. We followed this core concept and used Random Forest with features inspired by earlier studies. This approach achieved

73% accuracy on a data set of texts written by students in Cambridge English examinations (Xia et al., 2016). The classifier predicts five different levels of readability varying from A2 (easy) to C2 (difficult) (Xia et al., 2016). We map these values to percentages so that A2 becomes 100% (easy to read) and C2 becomes 20% (difficult to read) (see Table 1)

Table 1: Levels of readability

Text level	A2	B1	B2	C1	C2
Value	100%	80%	60%	40%	20%

2.4 Sentiment

A text containing sentiment is written in an emotional style. To determine the sentiment value of an article, our algorithm uses the pattern3.en library (Hayden and de Smet). In this library every word is assigned a sentiment value, which can be negative or positive [-1; 1]. If a word shows intense positive emotions (e.g. happy, amazing), it is given a high positive value. In line with that, a term indicating intense negative emotions (e.g. bad, disgusting) is assigned a high negative value. A word not containing any emotions (e.g. the, you, house), has a value of near to zero. First, the algorithm calculates the sentiment value for every sentence by averaging all absolute values of sentiment for the distinct words. After that, the overall sentiment value of the whole news article is calculated. For that, the average of the sentences is taken and multiplied by 100. [3]

2.5 Objectivity

Objectivity is given when a text is written from a neutral rather than a personal perspective. Phrases like *"in my opinion"* or *"I think"* are used by authors to reflect their individual thoughts, beliefs and attitudes. The process of determining the objectivity of a text is similar to the process of calculating the sentiment value. The aforementioned library pattern3.en (Hayden and de Smet) also includes a value of subjectivity for every word.[4] Therefore we use it to obtain an objectivity score for articles. Values range from 0 to 1, with a value

near to 0 indicating objectivity and a value near to 1 indicating subjectivity. The overall score for subjectivity contained in an article is calculated as the average over all sentences. However, since we want to examine the objectivity and not the subjectivity of a text, the values need to be inverted:

$$Objectivity = 1 - Subjectivity \qquad (2)$$

This score is normalized by multiplying it by 100 to attain a consistent score range for all labels.

2.6 Political bias

Bias measures the degree to which an article is written from a one-sided perspective that enforces users to believe in a specific viewpoint without considering opposing arguments.

For calculating political bias we followed Fairbanks et al. (2018) and used two classes that represent different political orientations: conservatism (sources that are biased towards the right) and liberalism (sources that are biased towards the left). The authors also argue that the content of the article is a strong discriminant to distinguish between biased and non-biased articles. Following the authors we built a *content based model* for prediction of political bias in the news articles. To achieve that, a logistic regression classifier is trained on a dataset containing articles from The Global Database of Events Language and Tone Project (The GDELT Project). This database monitors the world's broadcast news in over 100 languages and provides a computing platform. However, it does not contain any information about the political bias. To retrieve the bias contained in an article, we crawled from the Media Bias Fact Check [5] the required bias information. The Media Bias Fack Check contains human annotated fact checks for various source domains. For our articles we have left-biased, right-biased and neutral labels. We use a simple bag-of-words approach as features to guide our logistic regression model. As the label values in our plugin are all shown in a range from 0-100%, the label's landing page shows 0% when the article has no political bias otherwise 100% – regardless whether the article is left or right biased. When the label's card is flipped the reader can see whether the article has left or right political bias.

---

[3]Since we use the absolute values of the sentiment scores we are interested in knowing how sentimental a news article is rather than focusing on the valence of emotions.

[4]Similar to sentiment the algorithm calculates the subjectivity score for every sentence by averaging all subjectivity scores of its words.

[5]https://mediabiasfactcheck.com

3 Visualization

3.1 Colors and icons

Our information nutrition labels are represented by simple, easy-to-identify and well-known icons, which have been shown to be easily understood by users (Antunez et al., 2013; Campos et al., 2011; Hersey et al., 2013; Roberto and Khandpur, 2014). Moreover, previous work reports that simple additional texts allow for a quicker processing of information represented by an icon (Campos et al., 2011). Therefore the information nutrition label is shown additionally as text.

To make nutrition labels more comprehensible, colors indicating amounts of nutrients are helpful (Aschemann-Witzel et al., 2013; Crosetto et al., 2016; Ducrot et al., 2016). Relevant research reports that both traffic light colors and monochromic colors work equally well (Aschemann-Witzel et al., 2013). Traffic light colors are most common with red indicating high (i.e. negative) and green indicating low (i.e. positive) levels of nutrients (Kim et al., 2018). Since we do not want to bias the users towards reading an article or not, but rather give information about its content, we chose to use different shades of blue in our plugin. A light blue indicates low and a dark blue indicates high levels of a certain label. Additionally, blue stands for trust, honesty and security (Venngage, 2018), which should indicate that the user is operating a reliable tool.

When deciding on what charts and figures to use, we again took into account that simple and commonly known visualizations are easiest to comprehend (Campos et al., 2011). Thus, we chose plain bar charts for representing overall nutrition label scores as well as scores of sub-labels. Additionally, we enriched it with percentages as well as coloring. Consequently, the amount of a label contained in a news article is visualized in an understandable and easy-to-process way.

For lettering, the font *Futura* is used. It is a modern, straightforward and clean typeface often used in state-of-the-art websites and fits the simple and genuine layout.

3.2 Positioning and information distribution

Following the so-called gestalt laws of grouping[6], objects that are closer to each other (*law of proximity*) are perceived as belonging together. More-
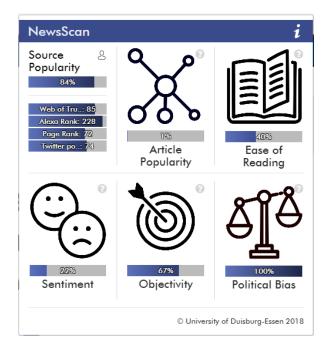
---

[6]e-teaching.org

Figure 1: NewsScan plugin

over, to indicate the grouping of information, separations between those groups are useful (*law of continuity*). Therefore, the distinct labels are separated from each other by horizontal and vertical lines.

To obviate information overload (Eppler and Mengis, 2004) when using *NewsScan*, we reduced the information on the landing page to a minimum. Only the icons with their respective overall score are visible on the front side of the cards we used for visualization (as shown in Figure 1 for article popularity, ease of reading, sentiment, objectivity and political bias). Therefore, the users can get a first impression about different nutrition labels of the article. When time is scarce, a simple visualization where users can find the demanded information easily is most practicable (Crosetto et al., 2016). However, if users are interested in getting more detailed information, we created a backside for each card. The backside also shows the total score and, if available, relevant sub-labels that are used to calculate the overall scores (see Figure 1: source popularity). Additionally, on hovering over the names of the labels and sub-labels, the user gets a short explanation about what the wording and score mean. To further avoid possible confusion on the user side, all of the labels are represented in the same way. Overall and sub-label scores are mapped to a range from 0-100% and

icons, texts and charts are arranged consistently.

4 Evaluation

To evaluate *NewsScan* in terms of wording, coloring and usability, we will conduct qualitative user studies. Participants will be interviewed and asked about their perception of the tool in general as well as concerning specific features. Since our aim is to not bias the user towards a consumption of one news article or another, we need to evaluate the plugin regarding that. Especially the wording of the features could affect the user in forming an opinion about an article. However, we want to help our users making an informed decision, so we need some kind of guiding, hence wording. To ensure that our tool only works as a guide and not a specific recommender, we do not interpret, for example, an easy-to-read article as being not worthwhile reading but just easy to understand. However, we believe some threshold label values about worthwhile and not worthwhile articles would indeed help readers in their decision making. In our evaluation we will aim to incorporate such information and draw conclusion between cases with threshold and without threshold values.

5 Conclusion and implications

In this paper we introduced *NewsScan*, a browser plugin to assist consumers of online news websites in their decision making about the content they engage with. Readers are guided to make an informed decision about editorials based on six labels: *source popularity*, *article popularity*, *ease of reading*, *sentiment*, *objectivity* and *political bias*. Label values are computed when a news article is retrieved. Through simple visualizations and an intuitive design, the user is confronted with the meta-information of the respective piece. To avoid biasing the user in any way with respect to the consumption of an article, the information is solely presented but not interpreted. Judgments about news sources should be made by users themselves. If an article is read or discarded relies on the user's opinion and individual weighing of the importance of the six labels.

In our immediate future work we plan to conduct user studies to analyse the validity of information nutrition labels and their usefulness for users. We also plan to investigate and integrate further information nutrition labels. Moreover it would be interesting to apply *NewsScan* to further media like videos or images accompanying news.

References

Lucia Antunez, Leticia Vidal, Alejandra Sapolinski, Ana Gimenez, Alejandro Maiche, and Gaston Ares. 2013. How do design features influence consumer attention when looking for nutritional information on food labels? results from an eye-tracking study on pan bread labels. *International Journal of Food Sciences and Nutrition*, 64(5):515–527.

Jessica Aschemann-Witzel, Klaus G. Grunert, Hans C.M. van Trijp, Svetlana Bialkova, Monique M. Raats, Charo Hodgkins, Grazyna Wasowicz-Kirylo, and Joerg Koenigstorfer. 2013. Effects of nutrition label format and product assortment on the healthfulness of food choice. *Appetite*, 71:63 – 74.

Sarah Campos, Juliana Doxey, and David Hammond. 2011. Nutrition labels on pre-packaged foods: a systematic review. *Public Health Nutrition*, 14(8):1496–1506.

Paolo Crosetto, Laurent Muller, and Bernard Ruffieux. 2016. Helping consumers with a front-of-pack label: Numbers or colors?: Experimental comparison between guideline daily amount and traffic light in a diet-building exercise. *Journal of Economic Psychology*, 55:30 – 50. Special issue on Food consumption behavior: Economic and psychological perspectives.

Pauline Ducrot, Chantal Julia, Caroline Mjean, Emmanuelle Kesse-Guyot, Mathilde Touvier, Lopold K. Fezeu, Serge Hercberg, and Sandrine Pneau. 2016. Impact of different front-of-pack nutrition labels on consumer purchasing intentions: A randomized controlled trial. *American Journal of Preventive Medicine*, 50(5):627 – 636.

Martin J. Eppler and Jeanne Mengis. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines. *The Information Society*, 20(5):325–344.

James Fairbanks, Fitch, Knauf, and Briscoe. 2018. Credibility assessment in the news: Do we need to read? *ACM ISBN 123-4567-24-567/08/06*.

Norbert Fuhr, Anastasia Giachanou, Gregory Grefenstette, Iryna Gurevych, Andreas Hanselowski, Kalervo Jarvelin, Rosie Jones, YiquN Liu, Josiane Mothe, Wolfgang Nejdl, et al. 2018. An information nutritional label for online documents. In *ACM SIGIR Forum*, volume 51, pages 46–66. ACM.

Tim Gollub, Martin Potthast, and Benno Stein. 2018. Shaping the information nutrition label. *ECIR*.

Andy Hayden and Tom de Smet. Pattern 3. Retrieved July 27, 2018 from https://github.com/pattern3.

James C Hersey, Kelly C Wohlgenant, Joanne E Arsenault, Katherine M Kosa, and Mary K Muth. 2013. Effects of front-of-package and shelf nutrition labeling systems on consumers. *Nutrition Reviews*, 71(1):1–14.

Eojina Kim, Liang (Rebecca) Tang, Chase Meusel, and Manjul Gupta. 2018. Optimization of menu-labeling formats to drive healthy dining: An eye tracking study. *International Journal of Hospitality Management*, 70:37 – 48.

Christina A. Roberto and Neha Khandpur. 2014. Improving the design of nutrition labels to promote healthier food choices and reasonable portion sizes. *International Journal of Obesity*, 38(S1):S25.

Sarah E Schwarm and Mari Ostendorf. 2005. Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 523–530. Association for Computational Linguistics.

The GDELT Project. Watching our world unfold. Retrieved July 27, 2018 from https://www.gdeltproject.org/.

Sowmya Vajjala and Detmar Meurers. 2013. On the applicability of readability models to web texts. In *Proceedings of the Second Workshop on Predicting and Improving Text Readability for Target Reader Populations*, pages 59–68.

Venngage. 2018. What marketers should know about the psychology of visual content. Retrieved July 25, 2018 from https://venngage.com/blog/marketing-psychology/.

Menglin Xia, Ekaterina Kochmar, and Ted Briscoe. 2016. Text readability assessment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 12–22.