# Shaping the Information Nutrition Label

Tim Gollub[1]  Martin Potthast[2]  Benno Stein[1]

[1]Bauhaus-Universität Weimar
tim.gollub@uni-weimar.de
benno.stein@uni-weimar.de

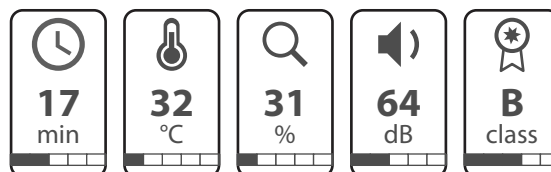[2]Leipzig University
martin.potthast@uni-leipzig.de

Figure 1: Visual representation of the proposed label. Time, temperature, transparency, volume, and credibility are taken as quantities to describe the nutrition facts of a document.

## Abstract

We take up on the idea of a nutrition facts label for online documents: the Information Nutrition Label. Such a label has the potential to increase the readers' ability to make an informed decision *before* the "consumption" of a news article or some other published online document. The basic ideas along with the dimensions (manifest, measurable text qualities, etc.) of such a label were proposed in [FGG+17]. The paper in hand focuses on the problem of an intuitive, unambiguous, and intelligible label presentation. For this purpose we (1) categorize the originally proposed information nutrition dimensions and (2) interpret them in terms of well-known physical quantities from which we belief that they are intuitively understandable for the general public. To give an impression of our ideas, a visual representation as well as the results of a preliminary crowd-sourcing study are presented.

## 1 Introduction

The World Wide Web is a great source for news. However, relying on online news does not come without difficulties both for the individual and for a society as a whole [BMA15]. With the web's sheer endless stream of news on virtually any topic, readers get easily trapped into filter bubbles and may become disconnected from important public discourses. Social networks stimulate the formation of echo chambers where groups of like-minded people share hyperpartisan news while being unaware of alternative arguments and opinions. Furthermore, part of the news articles are spread not only for the purpose of informing people but come with a commercial or a political incentive. Publishers are not loath to use exaggerated or misleading claims and promises in teasers or headlines in order to catch readers' attention (clickbaiting), making it hard to assess the (trust-) worthiness of an article ahead of reading. To improve this situation, the authors of [FGG+17] propose a so-called "information nutrition label" for online news. Like its food counterpart, the label is supposed to help people making more informed decisions upon which news items to consume.

Starting from a set of nine information nutrition dimensions that have been proposed in the original work (see Section 2), we now further shape these ideas towards fewer categories as well as an intuitively understandable representation of the underlying, often complex text analysis results (see Section 3). The proposed label is based on a categorization of the nine information nutrition dimensions into the five categories I Effort, II Kairos, III Logos, IV Pathos, and V Ethos. The interpretation of these categories shall be simplified if associating them with well-known quantities from physics (or finance in the case of V); Figure 1 shows a possible implementation of the idea. I.e., a text is first analyzed regarding the nine information nutrition dimensions whose resulting values are then combined and rescaled to match the interpretation of the five categories. In this regard, the value ranges that are employed for the proposed categories were determined with a small crowdsourcing study (see Section 4).

Table 1: Categorization of the information nutrition dimensions (Column 1) into five categories (Column 2). Column 3 and 4 show the related physical/financial quantities and the proposed value ranges respectively. Column 5 states prototypical user questions that the respective category addresses.

| Dimension | Category | Quantity | Range | Addressed User Question |
|---|---|---|---|---|
| Readability Technicality Verbosity* | I Effort | Time | 0 - 120 min | Does time allow the reading? |
| Topicality Virality | II Kairos | Temperature | 0 - 100 °C | Do others care? |
| Factuality Verifiability* | III Logos | Transparency | 0 - 100 % | How professional is the writing? |
| Emotion Opinion Controversy | IV Pathos | Sound pressure | 0 - 120 dB | Is the article subjective? |
| Authority Credibility Trust | V Ethos | Credit rating | class A+...D | How reliable is the source? |

## 2 Related Work

The idea of computing an information nutrition label is an outcome of the Dagstuhl Seminar "User Generated Content in Social Media", held in July 2017[1], and has recently been published as a SIGIR Forum article [FGG+17]. In their proposal, the authors suggest nine dimensions as attributes for the information nutrition label, with one dimension, "authority / credibility / trust", actually comprising three dimensions into one. Table 1 lists these dimensions in the left column; for a detailed description as well as for a related discussion we refer to the SIGIR forum article. Note that Table 1 introduces two additional information nutrition dimensions, namely "verbosity" and "verifiability", which we see as complementary to the originally proposed ones. While "verbosity" refers to the length of an article, "verifiability" refers to the extent to which an article provides pointers to resources that help to verify claims made [HVER15].

We presume that a label that displays the nine original dimensions (along with their respective statistical measurement units) will receive attention mainly from experts. In order to open the results of an intricate document analysis to the general public (cf. the motivation for a traffic light system to simplify the food nutrition label [foo12]), we ask the question whether a simpler, yet equally informative and hence preferable label can be derived by merging those dimensions that make pragmatically similar statements.

## 3 Categorizing Information Dimensions

When studying the original nutrition dimensions it becomes apparent that some of them, like topicality and virality, or emotion, opinion, and controversy, are similar from a pragmatic point of view. E.g., the topicality

and the virality of an article both represent temporal and sociological phenomenons, and both may be used to answer a question like "How much do others care about the article?". Because of correlations like these, we presume that the nutrition dimensions can be subsumed into categories without a significant pragmatic loss of information.

In the second column of Table 1, our proposal for such a categorization of the altogether 13 dimensions into five groups is shown. The labels for the categories II ... V have been chosen in accordance to Aristotle's modes of persuasion [AK07], including the less well known concept of "kairos", which stands for the "right, critical, or opportune moment".[2] Column 3 shows our proposal for a category interpretation in physical terms (financial in case of V), along with sensible value ranges in Column 4. In Column 5, a prototypical user question which is addressed by the respective category is stated. In the following, each category is discussed in detail.

Category I, effort, groups all dimensions that affect the time a reader has to allot to comprehend an article. Besides verbosity, which has been used already by media websites to provide an estimated article reading time (e.g. by Medium[3]), the dimensions readability and technicality fall into this category. To express effort, we consider time in minutes as an intuitive choice. The effort category allows readers to check whether they have enough time to read an article and to identify articles of a specific depth.

Category II, kairos, groups all dimensions that pertain to the trendiness, momentum, or hotness of an article or a topic, i.e., topicality and virality. As an existing attempt to provide this category, the velocity graph [Pet13] on the media website Mashable[4] can be

counted. As a quantity to express kairos, we consider the temperature in the range of $0 - 100°C$ as intuitive. The kairos category can bring articles to readers' attentions which would be "out of their bubble" otherwise.

Category III, logos, groups all dimensions that capture how well an author supports her claims with evidence, i.e., factuality and verifiability. As a quantity to express logos, we consider transparency in the range from 0 - 100% as intuitive. The logos category can help readers to assess the journalistic quality of an article up front.

Category IV, pathos, groups all dimensions that are related to subjectivity and discrepancies, i.e., emotion, opinion, and controversy. As a quantity to express pathos, we consider volume, measured as sound pressure, as intuitive. The pathos category can help readers creating awareness that communities sharing alternative arguments or opinions likely exist.

Category V, ethos, finally groups all dimensions related to the credibility of an author or publisher, i.e., authority, credibility, and trust. As a quantity to express ethos, we consider credit ratings in the range from A+ to D, as used in finance,[5] as an adequate choice. The ethos category can help readers assessing the risk of becoming misinformed or, alternatively, the potential of learning about non-mainstream viewpoints.

## 4 Discussion

We see three advantages when using the proposed categories as attributes for the envisaged information nutrition label instead of the original dimensions. First, the reduced number of attributes makes the label both easier to present and easier to digest in practical settings. Second, by resorting to well-known quantities for the categories, readers can intuitively interpret the label without the need of detailed instructions. Third, the chosen quantities allow for the design of a non-textual visualization of the nutrition label. On the other hand, the potential concerns should not be overlooked: first, the categorization may be not as lossless as anticipated, such that the five categories convey much less helpful information than do the original dimensions. Second, the quantities (or their visualizations) may lead to false intuitions about the document they belong to.

As a very first step towards clarifying some of these concerns, we have designed a mainly non-textual representation for our label (see Figure 1), which allows for a visual comparison with the tabular label presented in the SIGIR forum article. In Figure 1, each category is visualized by a rounded rectangle featuring a category symbol and an article-specific category value. For the latter, both the absolute value as well as its relative
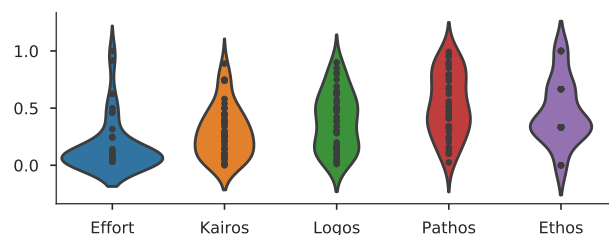


Figure 2: Violin plot for the annotations made by 42 Amazon Mechanical Turk workers for a news article from breitbart.com. The value range of all nutrition categories on the y-axis are scaled to the interval $[0, 1]$.

position in the value range are depicted. To make the values in the depicted label more sensible, we asked a crowd of 42 workers to read and then annotate the news article exemplified in the SIGIR forum article in the light of our categories.[6] As final value for the label we took the mean of all 42 annotations. For the sake of completeness, a violin plot showing the distribution of all annotations is shown in Figure 2. The violin plot indicates that the above article takes little effort to comprehend and, it is obviously not very hot anymore (kairos). In terms of transparency and sound pressure (logos and pathos), no clear consensus is reached, while the publisher is clearly not top rated in terms of credibility (ethos).

For future work, we consider to present a computational model for the information nutrition label and to further investigate the correlation of its constituent parts with human intuition.

## References

[AK07]  Aristotle and G.A. Kennedy. *On Rhetoric: A Theory of Civic Discourse.* Oxford Press, 2007.

[BMA15]  E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132, 2015.

[FGG+17]  N. Fuhr, A. Giachanou, G. Grefenstette, I. Gurevych, A. Hanselowski, K. Jarvelin, R. Jones, Y. Liu, J. Mothe, W. Nejdl, I. Peters, and B. Stein. An Information Nutritional Label for Online Documents. *SIGIR Forum*, 51(3):44–66, 2017.

[foo12]  foodwatch. Research supports traffic light colours. `https://www.foodwatch.org/more-information/research-supports-traffic-light-colours/`, 2012. Accessed: 2018-02-02.

[HVER15]  R. H. Harder, A. J. Velasco, M. S. Evans, and D. N. Rockmore. Measuring verifiability in online information. *CoRR*, abs/1509.05631, 2015.

[Pet13]  Robyn Peterson. Mashable launches google glass viral prediction app. `https://mashable.com/2013/05/14/mashable-launches-velocity-for-google-glass/`, 2013. Accessed: 2018-03-17.

---

[5] `https://en.wikipedia.org/wiki/Bond_credit_rating\#Credit_rating_tiers`

[6] `http://www.breitbart.com/big-government/2017/07/25/trumps-attack-on-sessions-over-clinton-prosecution-highlights-his-own-weak-stance/`