D4.2 Description of the Contropedia Platform

Due Date of Deliverable: M12
Actual Submission Date: 31/03/2015

Revision: Final

Start date of project: April 1st 2014    Duration: 12 months

Organization name of lead contractor for this deliverable:
Barcelona Media

Editor: David Laniado
Authors: David Laniado, Erik Borra, Michele Mauri, Andreas Kaltenbrunner, Esther Weltevrede, Paolo Ciucarelli, Giovanni Magni, Richard Rogers, Tommaso Venturini
## Project Information

<table>
<thead>
<tr>
<th><strong>PROJECT</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Open Call Project name:</strong></td>
<td>Real time analysis and visualization of controversies around Wikipedia articles.</td>
</tr>
<tr>
<td><strong>Open Call Project acronym:</strong></td>
<td>Contropedia</td>
</tr>
<tr>
<td><strong>Project start date:</strong></td>
<td>01/04/2014</td>
</tr>
<tr>
<td><strong>Project duration:</strong></td>
<td>12 months</td>
</tr>
<tr>
<td><strong>Contract number:</strong></td>
<td>288021</td>
</tr>
<tr>
<td><strong>Activity:</strong></td>
<td>THEME ICT-2011.1.1: Future Networks</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>DOCUMENT</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Document title:</strong></td>
<td>Description of the Contropedia platform.</td>
</tr>
<tr>
<td><strong>Document type:</strong></td>
<td>Report</td>
</tr>
<tr>
<td><strong>Deliverable number:</strong></td>
<td>D4.2</td>
</tr>
<tr>
<td><strong>Contractual date of delivery:</strong></td>
<td>Month 12, 31/03/2015</td>
</tr>
<tr>
<td><strong>Calendar date of delivery:</strong></td>
<td>Month 12, 31/03/2015</td>
</tr>
<tr>
<td><strong>Editor:</strong></td>
<td>David Laniado</td>
</tr>
<tr>
<td><strong>Authors:</strong></td>
<td>David Laniado, Erik Borra, Michele Mauri, Andreas Kaltenbrunner, Esther Weltevrede, Paolo Ciucarelli, Giovanni Magni, Richard Rogers, Tommaso Venturini</td>
</tr>
<tr>
<td><strong>Workpackage number:</strong></td>
<td>3</td>
</tr>
<tr>
<td><strong>Workpackage title:</strong></td>
<td>Integration</td>
</tr>
<tr>
<td><strong>Lead partner:</strong></td>
<td>Barcelona Media</td>
</tr>
<tr>
<td><strong>Dissemination level:</strong></td>
<td>CO</td>
</tr>
<tr>
<td><strong>Date created:</strong></td>
<td>01/03/2014</td>
</tr>
<tr>
<td><strong>Updated:</strong></td>
<td>31/03/2014</td>
</tr>
<tr>
<td><strong>Version:</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Total number of Pages</strong></td>
<td>22</td>
</tr>
<tr>
<td><strong>Document status:</strong></td>
<td>Final</td>
</tr>
</tbody>
</table>
Executive Summary

This document contains a description of the Contropedia platform in its current state.

After a short introduction on the aims of the platform, the document contains three main parts:

1. **Description of the Contropedia interface:** a detailed description of the platform’s interface and of the ways a user can interact with it to explore controversies within a Wikipedia article. This includes a description of all the views offered by the platform: the *layer view*, which provides an overlay over the original article, highlighting controversial elements; the *dashboard view*, which shows a ranking of the most controversial elements in the article, together with a timeline of when each element underwent more activity and dispute, and the users involved; the detailed view, which shows all the activity related to a given element. The section is meant as a kind of user’s guide.

2. **Description of the algorithms and metrics:** a description of the algorithms used to produce the visualizations and the metrics available in the platform. This includes the methods for extracting and processing the data, for computing controversiality scores based on the edit history, and for connecting discussion and edit activity by associating discussion threads to elements in the article.

3. **Directions for further development:** the last section describes our work in progress and our plans for improving and extending the platform.
Table of contents

1 Introduction .......................................................................................................................... 5

2 Description of the Interface (Contropedia user’s manual) ............................................... 6
   2.1 General layout .................................................................................................................. 6
   2.2 The article timeline .......................................................................................................... 6
   2.3 The layer view .................................................................................................................. 6
   2.4 The dashboard view ......................................................................................................... 7
   2.5 Detailed views .................................................................................................................. 9
   2.6 Menu and options .......................................................................................................... 12

3 Algorithms ............................................................................................................................ 15
   3.1 Pre-processing ................................................................................................................ 15
   3.2 Associating wiki links to edits ....................................................................................... 15
   3.3 Controversy score ......................................................................................................... 15
   3.4 Matching editing and discussion activity ....................................................................... 16

4 Directions for further development .................................................................................... 18
   4.1 Improving the interface ................................................................................................. 18
   4.2 Refining the algorithms ............................................................................................... 19
   4.3 Optimizing the implementation ................................................................................... 20
   4.4 Extending Contropedia ............................................................................................... 20

5 Conclusions .......................................................................................................................... 21

6 References ............................................................................................................................ 22
1 Introduction

Wikipedia articles are not static documents, as an inattentive reader might think, but living artifacts that are liable to be modified over and over again. They often become the battleground for competing points of view that need to be turned into a neutral description gathering community consensus.

In line with its policy of transparency, Wikipedia makes the complete edit history and discussion pages of every article publicly available. However, these logs can be huge, displaying only the meta information of thousands of revisions in chronological order, and are disconnected from the discussions on the corresponding talk pages. Identifying disputes and negotiations that led to the current version of an article may thus require a significant effort, especially for big or controversial articles.

This is an issue both for readers who want to know more about an article's development, and for editors who have to manually look through the edit history and talk pages to find out when and why something was changed.

The platform described in this document aims to make such valuable information easily accessible through a visual interface that allows one to identify the elements that aroused more dispute and activity, and to explore the development of a topic within an article.

The demo and the open-sourced code for our platform are available at http://contropedia.net

Two peer-reviewed papers about the platform have been accepted for presentation at international conferences [1,2]. We have another paper under submission in a top journal [6] and are furthermore writing another paper with all participants of the consortium.
2 Description of the Interface (Contropedia user’s manual)

Collaborative content creation inevitably reaches situations where different points of view lead to conflict. In Wikipedia, the free encyclopedia anyone may edit, disputes about content in controversial articles often reflect larger societal debates. While Wikipedia has a public edit history and discussion section for every article, the substance of these sections is difficult to phantom for Wikipedia users interested in the development of an article and in locating which topics were most controversial. Our platform, Contropedia, augments Wikipedia articles and gives insight into the development of controversial topics. Contropedia uses an efficient language agnostic measure based on the edit history that focuses on wiki links to easily identify which topics within a Wikipedia article have been most controversial and when [1, 2]. It effectively splits an edit (i.e. changes to the previous revision of an article) into changed sentences, thereby allowing the user of Contropedia to easily detect all the changes around a specific topic.

2.1 General layout

The platform offers two main views for exploring activity in an article. The views allow the user of the platform to identify controversial elements in an article, and to understand when, why and how something was controversial.

Controversiality in Contropedia is conveyed by colours: the hotter the colour of an element, the more controversial it is. Colour usage of is coherent over the whole platform: every time a controversial element is visualized, independently from the analyzed element, the same colour code is used.

The starting point for exploring an article through Contropedia is the menu, always accessible by clicking the icon on the top right of the screen. Through the menu users can select the article to be analyzed, and choose between the two main views (the layer view and the dashboard view). The menu also allows users to specify a number of other options, which are described in detail at the end of this section.

2.2 The article timeline

A timeline at the top of the page indicates the number of edits per month for the full edit history of the article, as shown in Figures 1 and 2. This visual element is always present and shows the user in which periods most activity took place in the construction of the article. Via the menu, it is possible to restrict the analysis to a specific time interval. In this way, all metrics and visualizations always reflect the activity in the specified period. By default, the full history of an article is taken into account.

2.3 The layer view

The first view, the layer view, adds a visual layer to a Wikipedia article, highlighting wiki links in function of their controversiality. In this way, controversial elements can be immediately located in the page. The hotter the colour, the more disputed took place around that element.

On the right, the layer view also contains a minified version of the article, as a scroll-bar, which gives an overview of the page and allows one to locate which sections were the locus of most negotiations.
The layer view is an access point for a more detailed inspection of the activity around a specific element: when the user clicks on any coloured link, the element is expanded and the corresponding *edits table* is shown (described below).

### 2.4 The dashboard view

The dashboard view, shown in Figure 2, presents a ranking of the most controversial elements in an article, and shows when each of them was disputed. By default, only the top 20 most controversial elements are shown, but this number can be changed through the menu.

While in the layer view one can only see the wiki links which are present in the current version of the article, the dashboard view also displays those that might have been deleted from the article. Such elements are struck through, e.g. "ocean" in Figure 2.
Figure 2: dashboard view for article “Global warming”.

Level

Next to each element, the dashboard shows its overall level of controversiality, computed for the selected period. This is conveyed by the controversiality score, and the corresponding colour.

Edit activity and controversiality

Next to each element’s controversiality level is a gray timeline that displays the number of edits involving the element over time. The timeline is represented as a horizon graph. Below the horizon graph, a thin coloured bar conveys when the element was most disputed. While the two measures are of course correlated, they are not the same: while the gray chart shows the overall number of edits on a given month, the coloured bar is based only the edits that are used to calculate the controversiality metrics (i.e. substantial disagreeing edits). Therefore, an element can have a lot of edits on a certain month, indicated by the gray horizon graph, but if activity is mostly due for example to vandalism and anti-vandalism edits, or to new content added through many edits that do not delete previous content, then the coloured bar will indicate a low controversiality - despite the high level of activity.

For a detailed inspection of all the activity related to an element in the dashboard view, it is possible to access the edits table (described below), by clicking on the element’s timeline.
Type

The type conveys what kind of element is shown. Currently we only display wiki links, but we have plans to extend this to other wiki elements such as images and references.

Users

A gray bar shows the number of users involved in making substantive disagreeing edits to sentences which contain the element. Clicking on this bar gives access to a more detailed view of these users and their social dynamics (the edit network, described below).

2.5 Detailed views

Both the layer view and the dashboard view allow a detailed view to inspect activity around a specific element: the edits table. In the layer view one can access it by clicking a coloured wiki link, in the dashboard view by clicking the timeline. The dashboard view also features an edit network, which can be accessed by clicking on the “Users” bar on the right.

The edits table

The edits table shows the edit activity around an element and allows for a fine grained analysis of the specific changes to sentences containing the element. We are currently also experimenting with displaying the comments from the talk page which mention the element.

It is good to keep in mind that our algorithms split edits (i.e. changes to the previous revision of an article) into a set of modified sentences. We do so in order for the user of Contropedia to easily detect all the changes around a specific wiki element or topic within an article.

By default the edits table only shows substantive disagreeing edits, on the basis of which controversiality scores are derived. However, different kinds of edits that are not substantive and disagreeing (such as inserts, vandalism edits, section changes, etcetera) can be included in the table by checking the corresponding options in the menu. More details about these options are given below, in the description of the menu.

By default edits (and comments) are presented in reverse chronological order: the most recent ones are shown on top. The order can be reversed to see the oldest edits on top, by checking the option “sort edits table chronologically” in the menu.

An example for element “list of scientists opposing global warming consensus” in the “Global warming” article is shown in Figure 3.
Figure 3: Edits table for element “list of scientists opposing global warming consensus” in the article “Global warming”.

Table columns

Each edit to a sentence including the element under analysis is shown as a separate line in the table. Each line has the following fields:

- **Revision**: displays the Wikipedia revision id, and links to the Wikipedia page which shows the changes introduced by the edit, with respect to the previous version of the article. In case of reverted edits, this field additionally shows the revision in which the edit was reverted.
- **Edit**: this field shows the specific changes to introduced by the edit, i.e. it shows the difference between the sentence containing the element, before and after the edit. Text added is shown in green, while text removed is shown in red. Unchanged text has no colour. For example, in Figure 3, the first entry displays that the sentence “Some scientists publicly ...” was changed to “Only a small minority of scientists ...”.
- **User**: the name of the editor who made the edit, with a link to the user’s page on Wikipedia.
- **Edit summary**: the comment added by the user when submitting the edit. This field usually contains an explanation of the edit (i.e., what was edited, and why).
- **Section**: the article’s section in which the modified sentence is located.
● **Type**: edits can be of four types, according to the way in which they are related to the element under analysis:
  ○ *Sentence change (s)*: a sentence including the element is edited
  ○ *Insert (i)*: the element is added
  ○ *Delete (d)*: the element is deleted
  ○ *Element change (e)*: the text in the element itself is changed. In the MediaWiki markup code, used in Wikipedia, links are added using the syntax: [[target_page | label]]. Therefore, we consider as element edits the edits that modify the text within the brackets (typically the label accompanying the link).

● **Time**: the date and time at which the edit was made.

### Element statistics

Above the edits table, some overall counts are presented, as shown in Figure 3.

- the overall number of substantive and disagreeing edits involving the element. These are the edits which are taken into account when determining the controversiality score of the element;
- the number of distinct users responsible for such edits;
- the number of revisions corresponding to these edits (this is not exactly the same as the number of edits, because one revision can include edits to more than one sentence involving an element);
- the quantity of edits broken down by type;
- the sections in which these edits are located;
- finally, also the number of reverts in which the element is involved is shown. Note that these are not necessarily substantive disagreeing edits, but the number includes also vandalism edits or other edits which by default are not shown in the table (e.g. inserts).

### Comments from discussion threads

It is also possible to include in the edits table comments from discussion threads mentioning the element under analysis. The way in which discussions are associated to elements in the article is described in Section 3.4.

If the option “*include comments from talk pages*” has been marked in the menu, the table will show both edits and comments, mixed together in (reverse) chronological order.

Each comment is shown as a row in the table, with the same fields as the edits:

- **Revision**: in the case of comments, this field indicates the discussion thread to which the comment belongs, and links to the thread in the talk page.
- **Edit**: this field shows the text of the comment.
- **User**: the username of the editor who wrote the comment, with a link to the user’s page on Wikipedia.
- **Section**: in the case of comments, this field has value “Talk”.
- **Time**: the date and time (in seconds) in which the edit was submitted.

### The edit network

The edit network shows the interactions between the users editing sentences including a specific element. It is accessible from the dashboard view, by clicking on an element’s users bar. An example is shown in Figure 4.

When an editor makes an edit to a sentence which contains the element under analysis, an edge connects the editor to the previous editor of the sentence. Edges are curved and drawn in clockwise direction. So, for a given node, a clockwise curve is an outgoing link, while counterclockwise curve is an incoming edge. When an interaction is reciprocated, two edges in both directions form a loop between two nodes.
The size of each node is proportional to the number of connections. One can thus see which editors are central to the development of the topic, such as William M. Connolley in the example shown in Figure 4, represented with a light blue node, a “network star” interacting with many users.

Colors are assigned to convey camps of opposing editors: while usual community detection algorithms group together nodes that have many connections with each other, and aim to represent homogeneous clusters of nodes, here we are doing somehow the opposite: the algorithm assumes that interactions represent disagreement, and assigns the same color to editors who do not interact much with each other, but interact with the same other editors. In this way, editors who have the same colour are editors who don’t delete content of each other, but who engage in disputes with the same other users. In Figure 4 it is possible to see for example that the users interacting with William M. Connolley have orange colour, while users interacting with orange nodes have mostly the same colour as William M. Connolley.

The algorithm, called Antagonist-Based Community Detection (ABCD-plugin), was developed during the first hackathon, and has been described in detail in deliverables D2.1 and D3.1, and is available on GitHub [3].

2.6 Menu and options

The menu, always accessible through the icon on the top right corner of the screen, allows users to select an article, and to define a number of options:
Figure 5: Menu with all the available options for the Contropedia interface.

- **Select view**: this option allows the user to choose between the two main views that Contropedia offers to explore an article, described in detail above:
  - the *layer view*, which provides an overlay over the original article, highlighting controversial elements;
  - the *dashboard view*, which shows a ranking of the most controversial elements in the article, a timeline of when each element underwent more activity and dispute, and the users involved.

- **Select article**: currently the platform is available for a sample of over 100 articles; this option presents all the available articles and allows selecting one through a drop-down list.

- **View settings**: this option allows the user to specify the number of elements shown in the dashboard view. By default only the top 20 most controversial elements are shown.

- **Date range**: this option allows the user to restrict the period of observation to a specific interval of time, by selecting a *Start* date and an *End* date. In this way, all the metrics and visualizations will be based only on the activity in the specified period. As it will be shown in the following, the article timeline at the top of the screen, as well as the actor timelines in the dashboard view may help the user to define a temporal interval of interest. By default, the whole history of the article is considered, until the last update of the data (reported below the article timeline).

- **Colour scale**: in Contropedia colours indicate the level of controversiality: the hotter the colour, the more controversial the corresponding elements. Colours can be assigned in two ways:
  - *capped* (default option): the colours are calculated based on absolute thresholds. This means that for example in a page that has undergone little controversy, even the most controversial elements will have a cold colour.
  - *relative*: all colours are distributed based on a normalization of the element with the highest controversiality score. This means that even in a page that has not had many disputes, the most controversial elements will always get the hottest colour.

- **Exclude users**: this option allows one to exclude some users from the data, by specifying user names separated by a comma. Such feature can be useful especially to filter out specific users such as bots.

- **Edits table extras**: These options change the (amount of) information which is included in the *edits table* view. These options do not change the calculation of the controversy score.
  - *sort edits table chronologically*: by default edits are sorted in reverse chronological order, i.e. most recent changes are shown first. By checking this option, the order of the changes will be reversed, i.e. the oldest changes will be shown on top.
  - *include comments from talk pages*: comments associated to an element will be shown, mixed with the edits in (reverse) chronological order. The techniques for matching comments to edits are described in the next section.

By default, Contropedia only shows substantive disagreeing changes to elements (which are used in the calculation of the controversy score). The next three options allow the user of Contropedia to see all edits to an element, even though such changes are not taken into account for the calculation of the controversy score.
○ **show section changes**: edits that add or remove an entire section are filtered out by default. By checking this option, such edits will be included in the edits table.

○ **show inserts**: checking this option will make *inserts* appear in the edits table. *Inserts* are defined as edits that add new content, without deleting previous content.

○ **show vandalism**: checking this option will also display vandalism edits in the edits table. Vandalism edits are mostly identified as such because they were reverted (i.e., the changes are undone by restoring the previous version of the article), and some element, such as the edit summary accompanying the revert, indicates that the reverted edit was vandalism. In such cases, both the vandalism edit and the corresponding revert are not used to calculate the controversy scores. With this option it is possible to see such edits in the edits table anyway.
3 Algorithms

While the previous section focused on the description of the interface, this section describes the back-end of Contropedia, i.e. the algorithms used to collect and analyze data from Wikipedia.

3.1 Pre-processing
We retrieve the full edit history for a controversial article, including the wiki text of each revision and the meta data conveying at what time it was edited, who the editor was, as well as the editor’s comment (edit summary). Whenever a user makes multiple consecutive edits, we only retain the last version made by the user and discard all intermediate edits. As we are looking for substantive disagreements, we discard vandalism edits and their reverts by identifying whether the comment of a revert contains the word ‘vandal’, whether the user name making the revert belongs to one of the known anti-vandalism bots, when an IP-edit is reverted within 60 seconds, or when the automatic edit summary (WP:AES)\(^1\) indicates that the content of a page was blanked or replaced by unrelated text such as curse words.

3.2 Associating wiki links to edits
We make use of Wikipedia’s MediaWiki markup to identify the most relevant elements in an article. The current version of the platform focuses on wiki links, as they identify the key concepts and entities of an article; they are the lenses through which we can look at the substance and activity of controversies within a Wikipedia article. Our approach seeks to associate edits to wiki links by taking the sentences in which these links reside as our basic unit of analysis.

Let \( \{ R_1, ..., R_{r-1}, R_r, R_{r+1}, ... \} \) be the set of revisions of a Wikipedia article. As we are specifically interested in disputes related to a wiki link, we consider the edit activity on a sentence level by comparing every revision \( R_{r-1} \) with its successor \( R_r \). We split each revision into sections and then pairwise compare corresponding sections of \( R_{r-1} \) with those of \( R_r \). If the text of the sections differs, we use a diff algorithm to identify the edited sentences, the exact changes made to them, and the wiki links they contain.

To further assure that edits to sentences containing wiki links convey disagreement, we discard edits where only insertions are made. We also discard full section inserts or deletes (as these are mostly due to renaming of sections). We thus only consider edits that are substantive: the revision is not marked as vandalism; and that show disagreement: the changes should (also) contain a deletion.

3.3 Controversy score
We are interested in finding out how controversial a wiki link \( W_k \) is and compare the substantive, disagreeing, edit activity of sentences in which \( W_k \) appears. Intuitively, the more wiki links appear in an edited sentence, the less focus there is on one particular wiki link. For every sentence \( S_j \) with a substantive disagreeing edit, the weight attributed to a wiki link is thus divided by the total number of wiki links \( w(S_j) \) that appear in that sentence.

A controversy score \( c(W_k) \) is assigned to every wiki link \( W_k \) that appears in a sentence \( S_j \) with a substantive, disagreeing edit, of a revision \( R_i \) (up to a given revision \( R_r \)) as follows:

\[
c(W_k) = \sum_{i=1}^{r} \sum_{S_j \in R_i} \frac{1}{w(S_j)}
\]

(1)

In other words: the number of sentences with substantive, disagreeing, edits that include \( W_k \) are summed over all revisions up to \( R_r \). In those revisions where the wiki link \( W_k \) appears in an edited sentence with other wiki links, the summand is divided by the number \( w(S_j) \) of links involved. A wiki link thus accumulates controversialness through counting and weighting the substantive, disagreeing, edits to the sentences in which it resides.

\(^{1}\) See http://en.wikipedia.org/wiki/Wikipedia:AES
As an example, consider Figure 1 where two substantive disagreeing edits are shown. The first sentence contains two wiki links\(^2\) (‘List of scientists opposing the mainstream scientific assessment of global warming’, and ‘scientific consensus’), and the second only one (‘List of scientists opposing the mainstream scientific assessment of global warming’). If we just take these two edits into account, the first link would get a controversy score of 1.5 (0.5 in the first edit and 1 in the second) and the second link would get a controversy score of 0.5.

To find out which wiki links are most controversial, i.e. around which wiki links most dispute took place, we simply rank the wiki links of the article by their overall controversy score.

Further details about this algorithm can be found in our paper [1].

### 3.4 Matching editing and discussion activity

Editing of a Wikipedia article and discussion in the talk page are two parallel activities for the construction of an article; they are complementary and strongly relate to each other. However, in Wikipedia there is no explicit connection between the two, which appear just as two separated interaction spaces. Presenting combined information from these two parallel activity streams is one of the key challenges for the Contropedia project.

In the first hackathon we worked at the integration of the discussion activity with the layer view, associating discussion threads to article sections, while in the second hackathon we focused on the integration with the dashboard view, matching activity in the talk page to specific elements in the article.

Currently, only the association of discussion threads to elements in the article is implemented in the front-end, to show comments related to a certain element in the edit view. Details about features that are not currently integrated (yet) in the interface can be found in previous deliverables: the techniques to match discussion threads to article sections are presented in deliverable D2.1, while controversy scores of elements based on discussion threads are described in deliverable D3.1. The following section describes the algorithms and features which are currently implemented.

**Matching discussion threads to article elements**

We treat each thread as a whole unit of content, which is matched to all the actors mentioned in the thread. However, we keep trace of which comments in the thread explicitly mention an actor, and whether the thread

\(^2\) In MediaWiki links are written as \([\text{anchor} | \text{anchor text}]\), i.e. \([\text{link to article} | \text{optional name of link in text}]\).
includes some actor in its title. In this way we can compute the strength of the matching for each comment.

The main way of connecting discussions in the talk page to activity in the article is based on text analysis. A discussion thread is associated to some element in the article if:

- its text contains the element’s name, or;
- some previous comment’s text contains the element’s name, or;
- the thread title contains the element’s name.

Furthermore, we consider the edit summary, i.e. the textual note left by a user while editing the article, to explain the edit. When a user explicitly mentions a discussion thread in the edit summary, the corresponding edit, as well as the elements involved, are associated to the discussion thread.

Finally, we introduced another criterion of matching discussion activity to elements in the article, based on co-occurrence of edits and comments by the same user in a short interval of time. When a user in a short time window (by default, 10 minutes) has both edited the article, and left a comment in the talk page, we assume that the corresponding comment and edit are connected.
4 Directions for further development

While we have already received many enthusiastic reactions about our platform by Wikipedians, scholars, and journalists, we are keen to further the development of Contropedia. In this section we summarize the main directions we envision Contropedia should go. Most of the features described in this section are already under active development. Although the EINS grant has come to an end, we are actively pursuing further grant possibilities to continue their development.

4.1 Improving the interface

Completing the integration of discussions from the talk pages

As described in the previous section, and in deliverables D2.1 and D3.1, we have already implemented modules for matching discussion threads from the talk pages to specific sections and elements in the article. We have also developed controversy metrics for elements and sections over time according to the associated discussion in the talk pages, considering the depth of a discussion thread and reply chains, i.e. mutual replies between users.

In the interface however the integration is not yet complete; in the next version of Contropedia, the interface will be enriched to account for discussions and their controversy, in the following ways:

- In the layer view, each section of an article will link to the discussion threads related to it, and exhibit the controversy of such threads. So, on the one hand the layer view will be complemented by showing the controversy of each section, based on discussion threads. On the other hand, it will be an entry point for inspecting discussion threads associated to the various sections.
- The dashboard view shows for each element a controversy level, a timeline and a users bar; all of these are currently based only on edit activity: they will be made to account for activity associated to each element in the talk pages too.
- The edits table will be improved: comments from discussion threads will be integrated in a visually more coherent way.
- The edit network will also account for activity in both the article’s edit history and the associated talk pages.

Multi-language visualizations

Our approach is language independent, so the platform can easily be used for the comparison of different language editions, as witnessed by the article on Homosexuality, already available for exploration in several language versions in the Contropedia demo.

Currently an MA student at Density Design (Milan, Italy) is writing her thesis on the inter-language visualization of Wikipedia articles with the aid of Contropedia. Such outputs will be used in our development of visualizations to allow users to easily compare what is controversial in different language editions about the same topic, as well as when the same elements were mostly disputed and why.
4.2 Refining the algorithms

Improving precision by considering subparts of a sentence

In the current version of Contropedia, an edit to a sentence is attributed to all the wiki links included in the sentence. In cases of long sentences containing many wiki links, this may lead to assign edits to elements that are far in the sentence, and not really related with the modified content. This issue is particularly important in the abstract, the section that typically undergoes most edit activity, and contains many wiki links.

To address this issue and improve the precision of the controversiality scores, we are refining the algorithm so that in case of sentences containing more than one wiki link, an edit to a sentence is only associated to the wiki links that are closest to the modified content.

Distinguishing different instances of the same element in a page

Currently the controversiality of each element is calculated by considering all the sentences including the element. In other words, it is the sum of the controversiality of all the difference occurrences of that element along the article. This is very good for the dashboard view, an abstraction which gives an overall ranking of the most controversial elements in an article; however, in the layer view, that indicates where dispute is located within an article, we are working to treat different occurrences of the same wiki link separately in different parts of the articles (e.g. by calculating controversiality scores for each element per section in the article).

Furthermore, within Wikipedia it is good practice to not overlink. This often means that often only the first occurrence of a concept is linked (e.g. “greenhouse gas”). However, that concept may recur later in the article, without being linked. We thus aim to include topics which are mentioned in an article, but not linked later in the text, for the calculation of the controversy.

Improving the matching of discussion threads to article sections and elements

Matching comments from the talk pages to content and edits in the article is one of the hardest challenges for the project. As there is no explicit association between the two spaces, we rely on a set of heuristics that we are continually testing and improving.

Further experiment with measures of authority or trustworthiness of an author

After a presentation of our tool to Jimmy Wales, the founder of Wikipedia, he said “I love the concept as a potential tool for both readers and editors” However, he personally also thought that the authority or trustworthiness of an author should be taken into account. We will thus continue to further experiment with our user-based measure as described in D2.1

Validation

Although various users have already indicated that our tool works well to highlight the most disputed elements and give insights in what is most controversial, we would like to make a more thorough evaluation of the various (types of) algorithms.

First, we could evaluate the tool in terms of benefits to the user. Here a two step evaluation would be in place: 1) locate what is controversial within an article without our tool, and 2) locate what is controversial in the page with our tool. We suspect that the speed gain and easy aggregation of edits to the elements will be hugely beneficial for our users.

Second, we could evaluate the accuracy and precision of the different metrics in our tool by comparing whether Contropedia correctly detects the controversies provided in the following sources of ground truth:

• Manually labelled controversial issues in an article, by 1) topic experts, 2) Wikipedia editors of the article.

4.3 Optimizing the implementation

Efficient data collection from the Wikimedia dumps

Currently data about each article is retrieved through the MediaWiki API. Adding an article to the platform thus implies a certain overhead for retrieving all the article revisions about it; the larger the article, the larger the cost.

We recently significantly improved the performance of this issue, by implementing a new parser that retrieves the data from full Wikipedia dumps - in which all information about all articles is present. In this way, we hope to easily scale up the platform to the size of the whole English Wikipedia, with much lower additional time needed for the addition of an article.

4.4 Extending Contropedia

Visualizing a network of articles

As described in deliverable D3.1, we are working on an interactive visualization of a network of articles, based on the hyperlinks between them. As we do not only know how controversial an article is, but also how controversial each link to another article is, the network could represent controversiality of articles, as well as of their relationships.

Such network visualization would allow the user to zoom out from a specific article, and explore the context around it. It could also be the default entry point for Contropedia, offering an overview on a wider topic, and allowing the user to zoom in on a specific article. One could thus follow (the controversiality of) wiki elements across different related articles.

Now that we have displayed the potential of focussing on the controversiality of concepts within an article, we are considering to also add another interface which is more directly centered on wiki-objects (rather than on pages). This could allow us to answer questions such as ‘in which articles is this concept disputed?’.

Exploring controversies within a set of articles

Currently Contropedia only allows for the exploration of an individual article. We aim to extend the platform by adding the possibility to select a set of articles and analyse and visualise the activity on all of them together.

Beyond wiki links: showing controversiality of other kinds elements

While currently the platform only considers wiki links as elements, i.e. focal points for exploration of activity, other kinds of elements, such as external references, images and templates can easily be treated in the same way.

Showing controversiality of each word

Contropedia relies on specific elements, wiki links, in an article, and leverages them as focal points to explore and aggregate activity. This approach is efficient and generally effective in showing users relevant content related to disputes about concepts and entities. However, some controversies might be not entirely captured by sentences including wiki links, and in some cases a more fine-grained analysis could be useful. To further increase the precision of Contropedia we envision another view, similar to the current layer view, which displays the controversiality of each word.
5 Conclusions

In this document we have provided a detailed description of the Wikipedia platform as it is in its current state. The interface description is a sort of detailed user’s guide to exploring controversies in Wikipedia articles through the Contropedia demo that is available online. All the methods, algorithms and metrics employed for producing such visualizations are described in the document. However, more possibilities have been explored, and more results produced along the project. All these further efforts, which are not integrated in the current version of the platform (and only referenced in this document), are included in previous deliverables of the project. We are actively going on in improving the algorithms and the visualizations, as described in the previous section which presents our main in-progress and envisioned next steps to extend Contropedia.

A version of the user’s guide manual will be made available on the Contropedia web site: http://contropedia.net and kept up to date so that it always describes the platform as it gets improved.
6 References


