

# Public Debates on the Web

Fabian Gilson<sup>1</sup>(✉), André Bittar<sup>2</sup>, and Pierre-Yves Schobbens<sup>1</sup>

<sup>1</sup> PRcISE Research Center, University of Namur, Namur, Belgium

{fabian.gilson,pierre-yves.schobbens}@unamur.be

<sup>2</sup> Centre for Natural Language Processing, Université catholique de Louvain,

Louvain-la-neuve, Belgium

andre.bittar@uclouvain.be

**Abstract.** With the advent of social media, any piece of information may be spread all over the world in no time. Furthermore, the vast number of available communication channels makes it difficult to cross-check information that has been (re-)published on different media in real time. In this context where people may express their positions on many subjects, as well as launching new open initiatives, the public needs a mean to gather and compare ideas and opinions in a structured manner. The present paper presents the **WebDEB** project, which aims to develop a collaborative platform where opinions, namely arguments, are gathered, analyzed and linked to one another via explicit relations. **WebDEB** relies on various Natural Language Processing modules to semi-automatically extract information from the web and propose meaningful visualizations to the platform's contributors. Furthermore, public actors may be identified and attached to the ideas they publish to create a structured knowledge base where annotated texts, extracted positions and alliances may be identified.

**Keywords:** Social web · Public debate · Argumentation · Speech acts · Natural language processing

## 1 Introduction

Today, social media has become a key platform for public debates. Presidential elections, financial scandals, document leaks and so forth are among the many subjects that are attracting public persons and individuals to express their opinions in real time. As more people jump into debate arenas, more opinions, facts or statistics are published by many sources. Due to this complexity, any attempt to sort through or structure this mass of information requires a significant summarization effort [2].

Many initiatives following *Wikipedia*<sup>1</sup>'s example have appeared in recent years allowing any person to create and record information, making it openly available across the web. Furthermore, one of the main focuses in the field of

---

<sup>1</sup> <https://wikipedia.org>.

Natural Language Processing (NLP) is the extraction and structuring of textual data, with an increasing focus on the social web [5, 8].

Public debates are a gold mine of ideas, facts and opinions, but extracting the main structures from such material is tough work for the human mind [2]. Furthermore, such a flow of information is difficult to channel and post-process by the public. To this end, we introduce **WebDEB** (**WD**), an openly accessible platform dedicated to structure and analyze arguments expressed publicly or retrieved from press articles. Centered on graphical representations of arguments, involved actors and the relations between them, **WD** is designed to help citizens in today's understanding of public debates. As an initial goal, the **WebDEB** platform is intended to offer an online facility to concentrate arguments gathered from many sources, introduced by contributors as well as automatically fed from various sources. As a second target audience, **WebDEB** is built as a pedagogical support in language teaching in the field of, among others, discourse analysis and reasoning for secondary and higher education students [7].

In the present paper, we first describe the current design of the platform, with a series of captures of its implementation in Sect. 2. We then discuss the state of validation and list current limitations as observed by the present community of users in Sect. 3. In Sect. 4, we look over related work in discourse analysis, argumentation theories and debate representations. We finally sum up the contribution of this paper in Sect. 5 and review future work on the conception and validation of the platform.

## 2 The **WebDEB** Platform

**WebDEB** is initially designed as a pedagogical platform for the teaching of sociological issues. It is also intended as an open platform to gather any form of debates. With industrial partners, **WD** has also been developed to offer support to journalists in their information cross-checking and archiving tasks. We first conceptually depict the structuring of arguments in **WD** and describe the NLP features we put in place to support users in their encoding tasks and to build rich visualizations. We also provide some captures of the current implementation.

### 2.1 Conceptual Description

The two main concepts in **WebDEB** are **Contributors** and **Contributions**, as shown in Fig. 1. A **Contributor** is any registered user that contributes to the **WD** database. A clear hierarchy of contributors has been defined for pedagogical purposes on the one hand, and for content monitoring on the other hand. A standard separation has been specified to distinguish **Groups** and **Permissions** in order to make user management flexible and extensible. Also, a particular attention has been paid to data access, integrity and visibility since **WD** is meant to be openly accessible, while holding sensitive data.

**Contributions** are any data inserted (and validated when necessary) by **Contributors**. Possible types of **Contributions** are **Actor**, **Text**, **Argument** and **Argument-Link**. **Actors** may be affiliated to other **Actors** with the aim of tracing associations

of Persons to Organizations, and also partnerships between Organizations. Actors may be involved in many ways in TextualContributions, e.g. as authors or publishers. TextualContributions have Topics associated to them for language processing, search and filtering purposes.

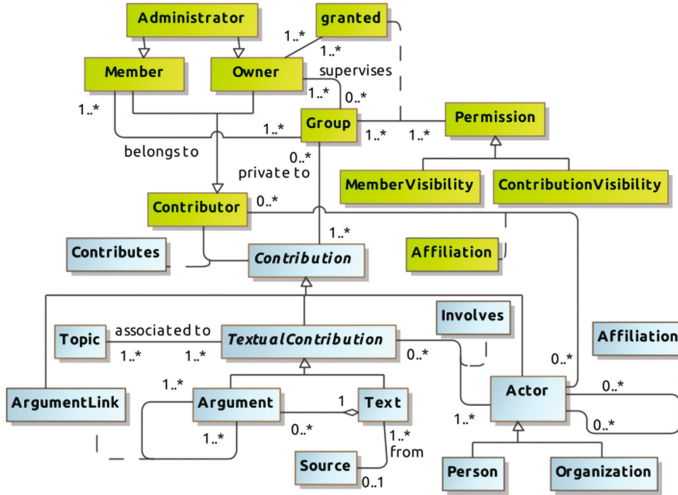


Fig. 1. WebDEB domain model

Arguments are extracted from Texts either manually or semi-automatically. Arguments are a rewriting of a text excerpt expressing a single idea in as simple form as possible. They also must be self-contained and unambiguous. Following the theory of *speech of acts* [1, 10], Arguments are classified into appreciative expressions of an emotion or a feeling, performative expressions of an action that changes the described situation, prescriptive expression of the necessity of taking an action or finally constative being any other type of expressions, like simple ascertainment. Temporal attributes (TimingType) and degree of certainty (ShadeType) are used to refine ArgumentTypes, adding more precision for NLP, searching and filtering purposes. Depending on the actual argument type, this degree is either specified on a three or five scale of belief in predefined textual form such as “I believe” or “It is likely that”.

Arguments may be interconnected to each other by justification links where one argument *supports*, *qualifies* (not really justifies but neither refutes) or *refutes* another one, or by similarity links where two arguments are *similar* (express the same point of view on a given subject), *nuanced* (somehow similar or dissimilar) or *dissimilar* (express opposing points of view on the same subject).

From ArgumentLinks between Arguments, accompanied by their authorships and affiliations of Actors, we are able to create valuable visualizations and aggregations that represent different viewpoints regarding the four aforementioned types of Contributions, as we will detail in Sect. 2.4.

## 2.2 Natural Language Processing and Data Retrieval Tools

A set of aids have been developed to assist contributors in encoding data or make suggestions based on semantic analysis of existing arguments.

*Argument classification.* Based on the aforementioned four types of *speech of act* [1, 10], a classifier described in a separated work is invoked to automatically suggest a classification when adding new arguments [3]. This classes are meant to describe the linguistic mechanisms used in the discourse and is used to discover argumentation links automatically. A second classifier is also invoked to retrieve the *discourse connectives*, e.g. comparison, condition or correlation.

*Automated data retrieval.* When adding press articles to the platform, web content may be imported with other properties such as the authors, simplifying the encoding of new texts that are saved either in the private library of a contributor, or when permitted by copyright, in the publicly visible library. Also, from partnerships with press editors, an *Rich Site Summary* (RSS) feeder regularly injects new content into the public database on which topic extraction processes are run. Another service listens to *Twitter* feeds and process them to create potential arguments that will be further validated by contributors. A final data retrieval service extracts actors' details, such as affiliations, from the *wikidata.org* open database based on a user-provided name or *Wikipedia* url.

*Similarity detection.* Last, when new arguments are inserted into the database, they are compared to all current ones in order to identify potentially similar arguments and to enrich the various visualizations presented in Sect. 2.4. Potential pairs of semantically similar arguments are then subject to manual validation by contributors before being effectively added into the WD platform.

## 2.3 Typical Usage

The core business of the WebDEB platform resides in arguments. In order to import them into the database, contributors either validate automatically extracted ones in a dedicated *yes/no* webpage where a list of suggestions is available, or they may start by importing a text like a press article. From any text, contributors may start to annotate them using the dedicated screen shown in Fig. 2.

Named entities, e.g. persons, organizations, dates or professions are highlighted from the text displayed in the top area with paging facility. For all arguments already extracted from the current text, their corresponding excerpts are highlighted too, giving the possibility to see the corresponding transformed arguments as well as going to their own visualizations.

Contributors are then able to select part of the text (with smart paging in case of an excerpt would be split onto multiple pages) to display an *add-new* pop-up window and fill in the details regarding that argument.

At the bottom of this screen, the list of already extracted arguments, the text structure in terms of these arguments, the properties and all involved actors are also viewable. The discourse structure is editable with a *drag and drop* feature, as visible in Fig. 3.

Hillary will:

Give working families a raise, and tax relief that helps them manage rising costs.

Create good-paying jobs and get pay rising by investing in infrastructure, clean energy, and scientific and medical research to strengthen our economy and growth .

Close corporate tax loopholes and make the most fortunate pay their fair share .

Hillary believes the defining economic challenge of our time is raising incomes for hardworking Americans.

Today's main economic challenge is to increase the incomes of hardworking Americans.

country is standing again, but we're not yet running the way we should be. From her first day in office to the last, Hillary Clinton will fight for you and for more take-home pay so you can get ahead and stay ahead.

Hillary understands that in order to raise incomes, we need strong growth, fair growth, and long-term growth. And she has a plan to get us there.

Strong growth

Hillary will invest in infrastructure, clean energy, and scientific and medical research to create jobs and strengthen our economy. And she'll provide tax relief to working families and small businesses. That's how we'll move toward a full employment economy that creates jobs, pushes businesses to compete over workers, and raises incomes.

Provide tax relief for families. Hillary will cut taxes for hard-working families to increase their take-home pay as they face rising costs from child care, health care, and sending their kids to college. She is calling for extending a tax cut of up to \$2,500 per student to help deal with college costs as part of her New College Compact, and for cutting taxes for businesses that share profits with their employees.

Highlighting : Organization Person Place Time Profession Connector

Arguments Justification links Text properties Text actors

Fig. 2. Annotated text, from which arguments are extracted

Arguments Justification links Text properties Text actors

You may edit justification links by dragging arguments from the left to the right and selecting the type of justification link you like (supports, qualifies or rejects).

Arguments Justifications

All arguments have been linked to each other

- Hillary Clinton promises to give working families a raise and tax relief that helps them manage rising costs.
- is supported by Hillary Clinton promises to close corporate tax loopholes and make the most fortunate pay their fair share.
- is supported by Hillary Clinton promises to create good-paying jobs and get pay rising by investing in infrastructure, clean energy, and scientific and medical research to strengthen our economy and growth.
- is supported by Today's main economic challenge is to increase the incomes of hardworking Americans.

Cancel save links Publish in group\_webdeb\_public

Fig. 3. Defining the argumentation structure

In order to add arguments, like other types of contributions, contributors are invited to fill in forms as presented in Fig. 4. The type of argument is proposed by the aforementioned annotator and the argument shades are proposed accordingly to the selected type. An automated topic extraction module has also been developed to help contributors when filling meta-data and suggestions of argument standardizations are currently under development.

When submitted, the argument will be highlighted in the text and compared to the existing database to find similarity matches to be validated later on.

Similar forms are available to encode the other types of contributions into the platform, but they are not detailed here for space reason.

Please fill in the properties regarding your argument. Fields marked with a \* are mandatory.

“ Hillary believes the defining economic challenge of our time is raising incomes for hardworking Americans. ”

Who enunciates the argument? \* ⓘ

Person or organization	Function (optional)	Affiliation (optional)		
Hillary Clinton	Function or mandate	Affiliation	+	-

Who reports the argument? ⓘ

Name	Function (optional)	Affiliation (optional)		
Person or organization	Function or mandate	Affiliation	+	-

Type of argument \* ⓘ

- Constative** It is a true fact, It is not true that,...
- Prescriptive** Something/ someone should/ must,...
- Opinion** I love, I hate, I feel,...
- Performative** I promise, I decide, I commit myself, I demand,...

Select one type from the available propositions

🔍 Our annotation service suggests the following types: Constative (100%)

Cancel

Fig. 4. First screen of the two step process to add a new argument

## 2.4 Rich Visualizations

For all types of Contributions, a set of visualizations have been created to trace arguments and explore related authors, texts and sources. Other visualizations have also been created to check opponents of Actors or summarize any argument made by an Actor that either has been encoded by users or automatically imported from external sources. Finally, a representation has been developed that presents the hierarchy of arguments inside a particular text, depicting its logical structure. On top of all these depictions, all Contributions, apart from the ArgumentLink which has no meaning *per se*, have a summary page where all their inherent properties are displayed<sup>2</sup>.

For Actors, a first view traces all affiliations and, for organizations, their affiliated Actors, as shown in Fig. 5. As in any other visualization, users are also invited to add more information. Furthermore, as for this view, graphs are exportable in many formats, *e.g.* PDF or PNG.

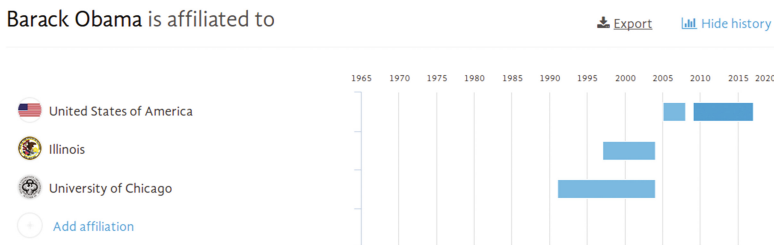


Fig. 5. Person's affiliations (historic view)

<sup>2</sup> These summary pages are not shown here due to space limitations.

All Arguments expressed by a given Actor may be displayed with an “*agreement graph*” depicting how many other Actors agree or not for each Argument this Actor has produced. This view may be sorted either by date, according to the amount of similar Arguments this particular Actor made, *i.e.* number of times this Actor said the same thing, or by the degree of agreement of other Actors, as visible in Fig. 6.



Fig. 6. All talks of a given Actor (sorted according to the degree of agreement)

For Actors, all their talks may be compared to the talks of all other Actors to build aggregated views of allies and opponents that may be sorted on an individual basis, or grouped by ages, functions, countries or affiliation (organization), as shown in Fig. 7.

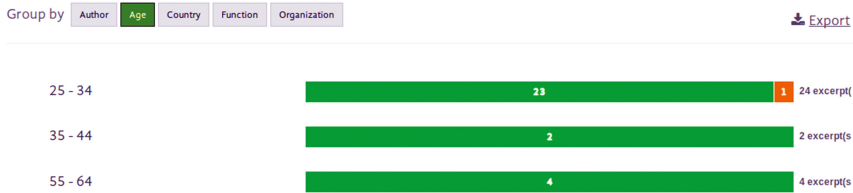


Fig. 7. All allies and opponents for a given Actor (grouped by ages)

The most significant visualizations regarding public debates are the similarity and justification maps of Arguments. First, the justification map, as presented in Fig. 8, articulates around a chosen Argument the other Arguments that support or refute it. We use transitivity rules over similarity relations to enhance justification maps. In the following rules, let:

- A, B, C be Arguments
- $\cong$  and  $\not\cong$  respectively denote *similarity* and *dissimilarity* between Arguments

–  $\Rightarrow$ ,  $\rightsquigarrow$  and  $\dashv$  respectively denote *justification*, *qualification* and *refutation* between Arguments.

We then apply the following transitivity rules (note that we do not use the *nuance* relationship to explore the similarity and they are only presented at the first *degree* of the justification map, as visible in Fig. 8):

$$\begin{array}{l}
 (A \simeq B) \wedge (B \rightarrow C) \quad \vee \quad (A \leq B) \wedge (B \dashv C) \quad \Rightarrow \quad A \rightarrow C \\
 (A \simeq B) \wedge (B \rightsquigarrow C) \quad \vee \quad (A \leq B) \wedge (B \rightsquigarrow C) \quad \Rightarrow \quad A \rightsquigarrow C \\
 (A \simeq B) \wedge (B \dashv C) \quad \vee \quad (A \leq B) \wedge (B \rightarrow C) \quad \Rightarrow \quad A \dashv C
 \end{array}$$

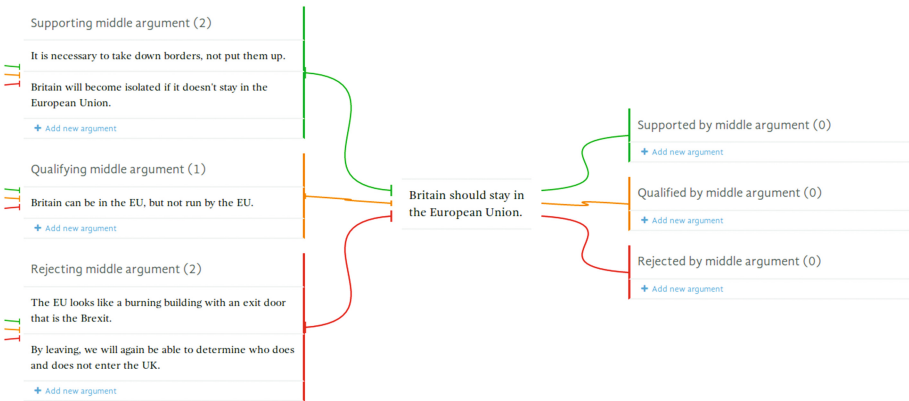


Fig. 8. Justification map of arguments

Similarity maps are displayed as sortable Argument lists where all Actors that have taken sides for or against a chosen Argument are shown, as presented in Fig. 9. Some statistics are also displayed to give a quick overview of the amount of Arguments having a similarity relationship with the Argument under investigation and the amount of Actors having partaken on the same subject. As for many other views, various grouping possibilities are available to the user.

### 2.5 Smart Search with Filters

Another reason of enforcing such structured details regarding contributions in general is to empower the ability to query the database and filter results in a user-friendly way. Since we are targeting a wide range of user profiles, from non-experts to journalists or sociologists, we made a point of providing an effective way of searching through the contributions. To this end, coupled to a common search bar, we added a filtering feature *à la* amazon, as shown in Fig. 10.





Fig. 9. Similarity map of arguments

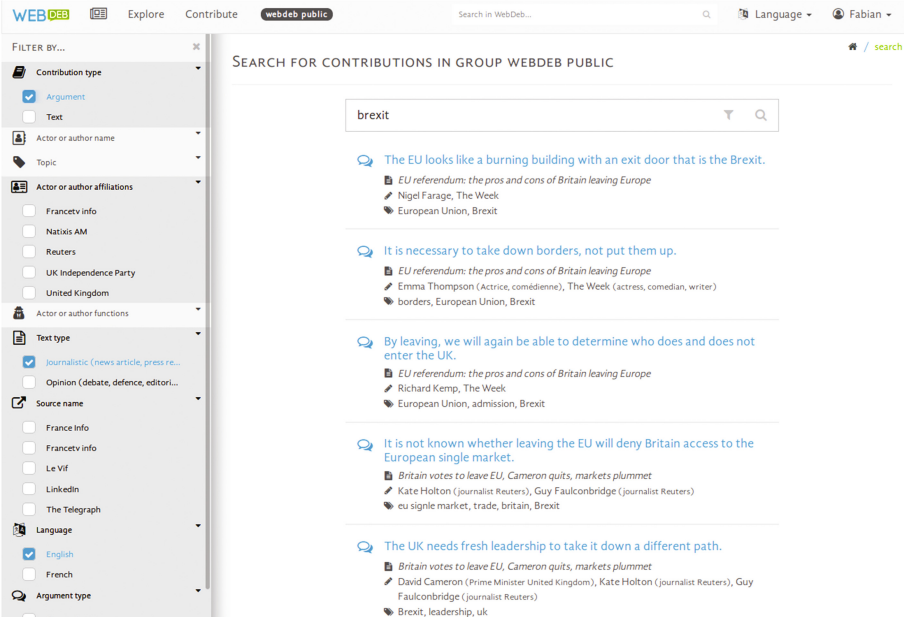


Fig. 10. Search and filter contributions

Filtering values on the left side are dynamically calculated based on the request made by the user. In the example shown, we may refine between all sources from which arguments have been extracted, the functions (professions) of the involved actors, their affiliations or their names, the topics, and so forth.

## 2.6 Group Management

As a pedagogic platform, WD also provides the possibility to create *closed* environments, named groups, where teachers may work with their students either collaboratively, or individually. Specific features have been put in place to this end such as the ability to activate or not the various NLP helps, the possibility to invite users to join groups, validate and mark students' contributions or to push validated contributions into the public database.

## 3 Discussion and Limitations

WebDEB provides dedicated visualizations to explore arguments expressed in the public arena. It is meant to provide graphical overviews of public actors with respect to particular topics or other actors' positions. Enhanced by NLP tools meant to facilitate the contributors' job while adding new data into the database, the data insertion burden that could lead to undesirable and discouraging effects, is minimized as much as possible.

We decided to work from a structured representation of arguments, texts and actors in order to ease the work of NLP-based discovery of relations between arguments as well as effectively tracing actor's positions. The argument's degree of certainty is meant to avoid formally linking arguments that have inappropriate levels of confidence. Thought, this *rigorous* structuring requires a couple of hours of familiarization to contributors before making things straightforward.

Content monitoring is also a rather harsh task, even if content may be deleted easily by (group) administrators. Improvements must be made regarding offensive words and fake news, at least by warning content administrators when detecting such contributions. However, since the platform is not intended to record individual's own opinions but dedicated to gather politicians or scientists' statements, such a risk is minimized, but must be still taken into account.

We conducted preliminary qualitative evaluations of the platform between May and November 2016. Researchers from the *Université Catholique de Louvain* and students of a secondary school in Namur were asked to evaluate the ergonomic and aesthetic aspects as well as the platform's effectiveness as a search engine and as a societal need. From these experiments, a list of limitations were identified from which a series of visualizations are still under development, *e.g. coalition of arguments* where all actors' positions are aggregated from similar arguments regarding a particular argument, or an aggregated view of texts that relate to each other through their respective linked arguments. Some NLP tools are also under development, especially as aids in encoding tasks regarding argument extraction and standardization, as well as automatic suggestions from press articles themselves.

## 4 Related Work

A series of approaches and applications have been proposed in the field of mind-mapping, public debates and discourse analysis. Thought, most of

these approaches concentrate on argument diagrams, mainly following Toulmin's [11] or Walton's [13] methods [9]. Among such approaches, *Compendium*<sup>3</sup>, *ThruthMapping*<sup>4</sup> and *Rationale*<sup>5</sup> focus on argumentation graphs, sometimes collaboratively. WebDEB, on the other hand, makes use of the theory of speech acts [1, 10] and focuses on the collection and aggregation of statements from various sources in order to build summary visualizations.

Some recent research focus on the detection of speech act types on web forums and emails [6] and in *Twitter* feeds [12, 14]. Contrary to this work, we use *Twitter* as a source from which to extract and transform (as well as classify) tweets to populate our database with arguments.

In the context of (public) debates, many didactic initiatives have emerged in recent years, such as the *International Debate Education Association* (idea)<sup>6</sup>, *ArgueHow*<sup>7</sup> or *CreateDebate*<sup>8</sup>, where students learn to argue about specific topics. Although we target a pedagogical purpose close to *idea*'s view, we also aim to concentrate actual public debates into a centralized database thus empowering individuals with an open social network where they are able to browse and visualize substantial data. *Debategraph*<sup>9</sup> is a very close system to ours where people may explore argumentation maps. However, on top of the representation of argumentation graphs, our purpose is to relate arguments to their authors and to build richer aggregations on arguments and actors. Also, we provide a search engine that is more user-friendly thanks to our filtering capabilities.

## 5 Conclusion and Future Work

We have presented WebDEB, a collaborative and open platform dedicated to structuring and tracing public debates as well as building valuable visualizations for end users. This project has a democratic objective of allowing people to search for public actors and review their positions over topics or regarding other actors. By automatically integrating sources from partner media, WD centralizes opinions and builds summarized representations of public debates, as well as gathering in one place as much information as possible regarding public statements and controversy. WebDEB also follows other existing platforms dedicated to teaching argumentation theory, but provides richer data visualization and makes use of Natural Language Processing aids.

At present time, a series of limitations have been pointed out by early testers, and some functionalities are still missing in our approach. In future, we plan to build more visualizations, especially regarding public actors and more exportation facilities are targeted to generate CSV files, aggregated PDF documents or

<sup>3</sup> <http://compendiuminstitute.net/>.

<sup>4</sup> <http://www.thruthmapping.com>.

<sup>5</sup> <http://www.reasoninglab.com>.

<sup>6</sup> <http://idebate.org> and <http://debatepedia.idebate.org>.

<sup>7</sup> <http://arguehow.com>.

<sup>8</sup> <http://www.createdebate.com>.

<sup>9</sup> <http://debategraph.org/>.

even *Argument Interchange Format* (AIF) [4]. We are also gathering more feedback from post-graduate students currently using the platform for educational purposes in a linguistic course. Last, we will investigate the possibility to analyze comments on *Facebook* the same way we are doing for *Twitter*. As an extend to our current approach, we plan to use replies and *retweets* to discover similar or opposed opinions of persons, and this way, enrich our visualizations.

**Acknowledgment.** This work is funded by the Walloon Region of Belgium, under convention no. 1318202 (*Programme Germaine Tillion*). The authors also thank, in alphabetic order, L.-A. Cougnon, B. Delvaux, C. Fairon, P. Francq, S. Roekhaut and D. Uygur for their various contributions to this project.

**Website and sources.** A running instance of the platform can be found at <https://webdeb.be>. The sources are distributed under the LGPL license and available at <https://bitbucket.org/fabgilson/webdeb-sources>.

## References

1. Austin, J.L.: How to Do Things with Words. Harvard University Press, Cambridge (1962). 2nd edn. (2005)
2. Bex, F., Lawrence, J., Snaith, M., Reed, C.: Implementing the argument web. *Commun. ACM* **56**(10), 66–73 (2013). <http://doi.acm.org/10.1145/2500891>
3. Bittar, A., Cougnon, A., Delvaux, B., Gilson, F., Fairon, C., Schobbens, P.Y.: Webdeb: learning to analyse public debate with the help of NLP. Submitted to *Traitement Automatique des Langues* on 28 October 2016. [https://staff.info.unamur.be/fgi/webdeb/TAL\\_en\\_ex.pdf](https://staff.info.unamur.be/fgi/webdeb/TAL_en_ex.pdf)
4. Chesñevar, C., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., Willmott, S.: Towards an argument interchange format. *Knowl. Eng. Rev.* **21**(4), 293–316 (2006)
5. Erhmann, M.: Les entités nommées, de la linguistique au TAL: statut théorique et méthodes de désambiguïsation. Ph.D. thesis, Université Paris 7, France (2008)
6. Jeong, M., Lin, C.Y., Lee, G.G.: Semi-supervised speech act recognition in emails and forums. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009*, vol. 3, pp. 1250–1259. Association for Computational Linguistics, Stroudsburg (2009). <http://dl.acm.org/citation.cfm?id=1699648.1699671>
7. Litman, D.: Natural language processing for enhancing teaching and learning. In: *Proceedings 30th Conference on Artificial Intelligence (AAAI)*, pp. 4170–4176. Phoenix (AZ), USA (2016)
8. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. In: *Lingvisticae Investigationes*, vol. 30, pp. 3–26. John Benjamins Publishing Company (2007)
9. Schneider, J., Groza, T., Passant, A.: A review of argumentation for the social semantic web. *Semant. web* **4**(2), 159–218 (2013). <http://dl.acm.org/citation.cfm?id=2590215.2590218>
10. Searle, J.R.: *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge (1969)
11. Toulmin, S., Rieke, R., Janik, A.: *An Introduction to Reasoning*, 2nd edn. Macmillian Publishing Co., Inc., New York (1984)

12. Vosoughi, S., Roy, D.: Tweet acts: a speech act classifier for twitter. In: Proceedings of the 10th International Conference on Web and Social Media, pp. 711–715 (2016). <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13171>
13. Walton, D.: Fundamentals of Critical Argumentation. Critical Reasoning and Argumentation. Cambridge University Press, Cambridge (2005)
14. Zhang, R., Gao, D., Li, W.: What are tweeters doing: recognizing speech acts in twitter. In: Proceedings of the 5th AAAI Conference on Analyzing Microtext, AAAIWS 2011-05, pp. 86–91. AAAI Press (2011). <http://dl.acm.org/citation.cfm?id=2908630.2908645>