

# CLUSTERING OF TEXTUAL INPUTS IN LARGE EPARTICIPATION PROJECTS

Alexander Prosser<sup>1</sup> and Cyril Velikanov<sup>2</sup>

## **Abstract**

*eParticipation projects often comprise deliberation among participants, where proposals are collected, discussed and rated to be processed in further stages of the eParticipation. In the case of large-scale audiences, this initial stage involves mass-online deliberation (MOD), which has to cope with a potentially very large number of proposals advanced by the participants. To enable clustering, MOD rely on human- (i.e. participant)-based appraisals of proposals given in the course of the participation project. Based on these appraisals, this contribution then proposes a clustering algorithm that makes use of the whole set of the above individual ratings. (Dis-)approval ratings are first weighted by the indication of clarity, that is, the higher the clarity rating assigned by a person, the higher the weight with which the (dis)approval rating will enter the clustering.*

## **1. Introduction**

### **1.1 The Issue**

Present day ICT enables better inclusion of citizens in public decision making. It is the ubiquity and 24/7 availability of the digital media that greatly enhances and promotes citizen participation, by facilitating their participation from abroad, or by accommodating working people who do not have time during the day, or by simply attracting the younger generation that is used to the digital media. However, many eParticipation projects have suffered from low participation (see the reviews in [1] and [2]), which, of course, undermines the political validity of the results of the participation process and induces political representatives to ignore the results thereof, which in turn invariably leads to disenfranchisement among the (potential or real) participants thereby creating a vicious circle.

If the authors of this contribution refer to large-scale or mass (online) deliberation, the notion does not refer to schemes selecting a sample of participants out of the entire population and holding a deliberative process with them, with the most notable example being the process proposed by J. Fishkin (“deliberative polling” [3], [4]). In another publication, the authors have already extensively argued why such approaches from a purely practical purpose do not lead to optimal inclusion of people and furthermore, why they are also problematic from a socio-philosophical point of view [5]. If, in the following, the authors therefore speak of mass online deliberation (MOD), they refer to schemes where all members of the relevant population have the opportunity to directly participate in the process.

Only a few digital participation projects in Europe can be called “mass” online deliberation projects, most notably the deliberation on the new Police Law in Russia [6], the newly-drafted Icelandic constitution [7], [8], the amendment to the Estonian constitution [9] and some proposed

---

<sup>1</sup> University of Economics and Business, Vienna, Welthandelsplatz 1, 1020 Vienna

<sup>2</sup> Cyril Velikanov, Memorial Society, Moscow, Malyy Karetnyy pereulok, No. 12, Moscow

legislation in Finland on off-road driving [10]. An extended analysis of these participation projects can be found in [5]; briefly summarizing, their common denominator was

- A large number of participants with thousands/tens of thousands of participants and contributions;
- They were all triggered and (largely) run by some branch of government; apart from Russia the countries were small to medium-sized;
- They were heavily advertised and received extensive media coverage;
- They had difficulties coping with the large number of contributions, which were largely overlapping and were the discussion diffused into different strands with a large amount of duplication; in some cases huge manual resources were mobilised to structure these strands, which is, of course, not a generally applicable solution – even for processes run by or with the government;
- Results of the deliberation process were either not implemented or diluted in implementation creating some disenfranchisement.

In all these cases, it became clear that conventional deliberation methods cannot cope with large-scale deliberation (arguably the rationale for Fishkin's selective approach). Therefore a new general model of deliberation has to be found, that allows fair and inclusive, yet structured and goal-oriented deliberation in very large groups; the authors have proposed such a general model in [11], which will be presented in the following section. However, the general model still needs refinement in terms of a "technical blueprint", where one aspect – that of clustering proposals – will be dealt with in this contribution.

## 1.2 The General Model

Considering the above examples, the following phases can be distinguished in virtually all these projects [11]:

- There were some preparatory actions, appointing experts, creating in some instances the legal framework etc. (Phase 0); this also involves the registration process, where either pre-existing logins are used or participants are registered in some other way;
- The *ideation* phase, that is collection of proposals and their commenting by participants who also grade or rank the proposals (Phase 1); in this phase, contributions are grouped into thematic clusters, which happens as an ongoing background job of the system;
- The *consolidation* phase (Phase 2), where the proposals in a cluster are consolidated into a single lead proposal by a team of editors; there are several ways of choosing the members of the consolidation team, one would be to invite the originators of the best-ranked proposal in the thematic cluster;
- The *editing* phase (Phase 3), where clusters are merged to produce a smaller number of competing lead proposals; this phase is supported by the system, nevertheless the main work is done by a team of editors merging the lead proposals of one or several similar clusters;

mechanics must be found to (i) select the editing team and (ii) to provide a decision process for the merger outcome and also to allow for dissenting opinions;

- The selection of one of these competing lead proposals either by the participants themselves or by an external body, for instance a national parliament (Phase 4).

This paper deals with the mechanics of the clustering phase based on a set of ranked proposals that are the outcome of Phase 1.

### 1.3 Why Not Facebook?

A very popular question in this context is the usage of readily available social media with Facebook being only a *pars pro toto* in this regard. There are several reasons, why we would be extremely sceptical about their usage:

- These media are typically under the jurisdiction of the U.S., where legislation on some key points (hate speech, erasure of statements qualified as hate comments<sup>3</sup>, data protection<sup>4</sup>) may differ markedly from European standards;
- In the past, Facebook has shown a rather arbitrary policy of shutting down pages for no apparent reason: The disappearance of the page of the City of Munich [16] and H.C. Strache (an Austrian politician, see [17]) may serve as examples.
- Probably the most important issue is the linear character of those media, which in our view does not lend itself to complex deliberation. For instance, the assessment and clustering mechanism which we consider instrumental in achieving a well-designed MOD can hardly be imagined within the framework of the general social media.

## 2. Assessment of Proposals

### 2.1 General Aims

The clustering is based on the proposals made in Phase 1, where every participant has the opportunity to (i) make proposals, (ii) to comment on them and (iii) to rank them. The general mechanics that can be observed in fora virtually everywhere, however, has some serious shortcomings:<sup>5</sup>

---

<sup>3</sup> An extreme example would be The National Socialist Party of America vs. Skokie (II) [12], [13] where a public rally by an openly racist and anti-Semitic organisation was permitted by the Supreme Court on the grounds of the First Amendment. In this regard, plans announced by European politicians (for an example see [14]) to “punish Facebook with fines of up to 50 million Euros per case” if they do not erase comments fast enough, which (by European standards) are objectionable, appear completely unrealistic: Facebook & Co. are subject to U.S. Law – and nothing else. A European subsidiary that may handle advertising clients from Europe can be shut down fast and the clients handled from the U.S. via the Internet.

<sup>4</sup> This does not only refer to looser standards for the commercialisation of data, but also some legal obligation to provide data on request of the authorities. [15]

<sup>5</sup> In the following, we will only consider proposals and comments/grades pertaining to proposals. The structure of such a deliberation system could, of course, be a lot more complex with counter and spin-off proposals emanating from discussing an initial proposal, facts amended to proposals etc. In such cases, there would be a full-fledged Issue-Based Information System (IBIS), cf. [19]; this would however not change the fundamental argumentation in the following sections.

- It enables “cliques” or mobbing in that organised or ad hoc groups systematically favour or denigrate proposals;
- Minority voices will therefore not get an adequate chance;
- Astroturfing [18] will get a chance for the same reasons;
- Late comers will not have the same chances as early proposals, as the latter would have already received a lot of attention and favourable (claque?) votes; any ranking based on such votes will create a systematic bias against these later comers;
- Commentators and those who rank the proposal will probably be influenced by the comments and grades already given and will be likely to “join the bandwagon”.
- The whole process will therefore not yield the best goal.

This may lead to a biased, limited discussion that is distorted by cliques and possibly astroturfed and that suppresses minority voices. The authors therefore propose a review process on proposals, where the reviewers (participants) are invited to review the proposal; all reviews (and possible comments) are only shown after the reviews have been collected. [11] This ensures an unbiased review and rules out actions by cliques or other organised groups. The exact selection process for the reviewers can differ and would have to be tested in pilot applications; possibilities include:

- A completely randomised selection;
- Weighted randomisation with the weights constructed according to “points”, which are awarded for cooperative and responsible behaviour, such as doing one’s reviews, not making comments marked as offensive, etc.;
- Some other algorithm linking reviewers to the area and nature of the proposal; however, we maintain that even then a random element should stay with the reviewer selection.

The aim is to minimise the usage of administrative staff and moderators and to provide a high degree of self-organisation in the process. However, this proposed general review process must be “economic” (for the reviewers) and easy-to-handle in order to get the necessary number of (invited) proposals. The following section suggests a scheme to achieve these aims.

## 2.2 Graphical Ranking Mechanics

The main idea is to structure the act of ranking into three stages: Clarity – Agreement – Assessment. In the following, consider Figure 1, which proposes a software-based “virtual joystick” to be used for quick, easy and yet rich assessment that produces parameters that can be used in the clustering of proposals as presented in Section 3.<sup>6</sup>

The fundamental question is certainly, whether the proposal to be ranked was understood. This is achieved in the left-most state depicted in Figure 1. The example shows a high degree of understanding. This indication is the starting point for indicating agreement with the proposal.

---

<sup>6</sup> For an earlier version, see [22].

Agreement has to be weighted by the degree of understanding, which appears to be intuitively obvious (eg, full agreement with a proposal one does not understand is worthless). Therefore, in the second step in Figure 1, the agreement triangle opens depending on the indicated degree to which one understood the proposal. Higher understanding thereby confers a higher opportunity to rate the proposal. Indication can then be given along a horizontal axis, whereby the user interface guides the user in his/her inputs. In the example, a rather high degree of agreement is indicated.

There remains yet a third element of rating – the quality assessment. Let us consider the rationale for this third element using an example: One may understand what is meant in a proposal, one may agree to it – and yet find it poorly worded. Hence, all three elements have to be included. Nevertheless, there are some issues to be solved:

- Subjectivity in assessment: Somebody used to high-level literary language may grade down a factually written proposal that to most of the audience would seem well-worded; hence, the benchmark for “well worded” is a moving one from reviewer to reviewer;
- It works the same the other way round: The contribution of a certain Mr. Wittgenstein may seem odd to somebody who gets his/her world-view from the yellow press;

Who gets to the third stage depending on the grading of clarity and agreement? This may concern two aspects:

- Does assessment of a proposal require a minimum degree of understanding? We would subscribe to that notion, as *any* further analysis requires a basic understanding of the proposal;
- Are reviewers who disagree (strongly) with the proposal allowed to quality-rate it? We would object to that as two issues can be predicted in this regard: (i) proposals one does not agree with would probably be systematically downgraded in an assessment by many reviewers, and (ii) the process becomes vulnerable to some “nasty” strategies, such as giving high quality marks for poorly worded proposals one does not agree with and vice versa thereby corrupting the “opponents” argumentation base.

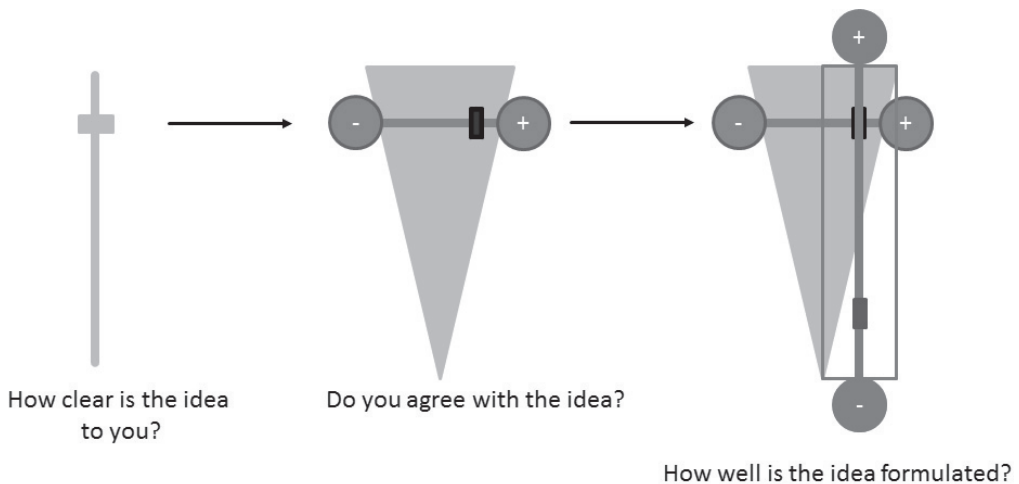


Figure 1: Assessment “joystick”

The mechanism proposed herein is subject to pilot implementations, which will provide a better understanding of the mechanism proposed on Figure 1 and its acceptance by real-world participants in a deliberation project. Therefore, the mechanism proposed can be seen as a falsifiable hypothesis that should be tested in a pilot implementation.

### 2.3 Implementation Considerations

Implementation-wise, this functionality could be realised in several ways, such as a Java Applet, Flash or HTML5 (of course in conjunction with java script). The most flexibility would probably be a Java Applet, which offers rich graphics and data handling behind the graphics; but this requires the download and installation of Java on the respective PC and may also create issues on mobile platforms. HTML5/java script would probably be the most general solution.

In any case, accessibility is an additional issue that needs to be addressed. Apart from the fairness and inclusion issues, public – and even private web sites in some countries – are by now required to offer accessible services. [20, 21] A possible solution would be to verbally request the ranking from the user in numeric form, incorporated in a dialogue that could run like the following:

*“How would you rate the clarity of the proposal on a scale from 0 to 10 with 10 being the maximum clarity?”*

*Response: 7*

*“How would you indicate your agreement on a scale between -7 and +7 where minus values indicate disagreement?”*

*Response: -4*

*“Thank you for grading this proposal”.<sup>7</sup>*

### 2.4 Outcome

The outcome of this process is a data set linked both to the reviewer’s account in the deliberation system and to the proposal, and containing a data triplet of (i) clarity grade C, (ii) degree of agreement (A) and (iii) quality grade (Q), providing the values for C and A exceed a certain threshold value. Once a sufficient number of grades has been collected, the general comment function is released by the system.<sup>8</sup> This is an important feature: The invited/appointed reviewers should do their work uninfluenced from either the ideas of the other reviewers as well as the general opinion (or the claue that professes to be “general opinion”).

This data triplet is now fed into the clustering algorithm that runs as a constant background activity during Phase 1.

---

<sup>7</sup> There is no quality assessment as the agreement rate is negative.

<sup>8</sup> Provisions have to be made if there is not a sufficient number of reviews, reviewers do not respond, etc. Also, this must be fine-tuned and tested in a pilot system.

### 3. Clustering

#### 3.1 Rationale for Clustering

In an MOD process, thousands and maybe even tens of thousands of proposals will be submitted. It stands to reason that a participant cannot read all of them, and that many – indeed most – of the proposals will be simply a duplication of others, where one and the same idea will be proposed over and over again. Therefore deliberating community needs to be provided with a “bird’s eye view” on the whole spectrum of proposals already submitted, where proposals are grouped and categorised for better reading and understanding.

This is done in the clustering process, which identifies highly similar or identical proposals, and groups them into a cluster. One of the proposals in a cluster is to become the lead proposal in the cluster, to some extent representing its common idea. In Phase 3 of the entire MOD model (see Section 1.2) these clusters, which may be numerous, are consolidated into a small number of distinctly different clusters. The clustering method should rely on the ranking method presented in the last section; user interaction that consists in providing explicit pair-wise comparisons of proposals is possible, but should be kept to a minimum. Such requests would run along the same principles as the reviews – that is, on invitation only and by randomly selected participants (or participants selected by some algorithm which includes an element of randomness).

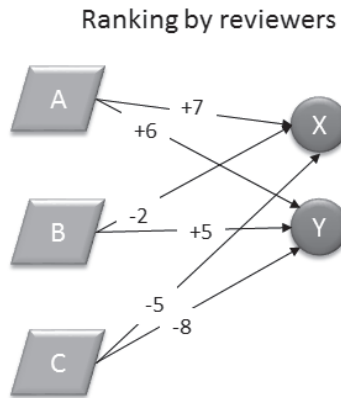
Clustering should be performed on a virtually constant basis providing a continuous update of the clusters as additional proposals and rankings thereof are submitted. Technically, clustering should therefore be implemented as a background job or at least a scheduled job running overnight.

#### 3.2 Method – Initial Clustering

Consider Figure 2. There are two proposals, X and Y; and three users (reviewers) A – C, who rank these proposals. Depending on their rating of their understanding of the proposals, they were able to rank their agreement to the proposals in a more or less pronounced way (cf. the discussion around Figure 1). A indicates very similar rankings for X and Y, the absolute difference being only  $D_A^{X-Y} = |7 - 6| = 1$ . C on the other hand dislikes both proposals,  $D_C^{X-Y} = |-8 - (-5)| = 3$ . The

fact that A likes and C dislikes the proposal is of no import at this stage. What counts is the difference between each reviewer’s rating. B has a different view. B supports Y, but not X, that is  $D_B^{X-Y} = |5 - (-2)| = 7$ . On average the difference in agreement ranking is 3.67 or generally

$D_i^{X-Y} = \frac{\sum |X_i - Y_i|}{n(i)}$  with  $X_i$  being the rating of proposal X by reviewer i.



**Figure 2: Distance measures in clustering**

There are two issues with this initial measure:

- While a proximity in agreement most likely signifies proximity in ideas, the same cannot be said about disagreement, one may, for instance, agree with a moderate idea to solve an issue and equally disagree with both extreme ways to address the issue. Hence, negative agreement ratings of both ideas should be dropped from the calculation of the average.
- Mixed positive/negative rankings such as that of B yield a high “spread” and tend to push the proposal in different clusters. However, we can learn something from that:
  - o If several reviewers rank like B and at the same time other reviewers rank the other way round, we may infer that X and Y almost certainly belong to different clusters;
  - o If at the same time there are reviewers who both support X and Y, we may infer that these proposals follow the same general line but differ in some aspects, probably one being more radical than the other.

In the last case, the issue is that a large number of reviewers supporting both proposals (like A) may outnumber the mixed reviewers (like B) and hence the proposals would be placed in the same initial cluster, where they most likely do not belong.<sup>9</sup> A way out of this bind may be to give mixed reviews like that of B a different weight from uniformly positive reviews like that of A.

A special case will be proposal pairs X, Y, where there are a lot of reviewers rating both proposals in a positive way  $\langle +X, +Y \rangle$ , some reviewers rating  $\langle +X, -Y \rangle$  but no mixed rating the other way round  $\langle -X, +Y \rangle$ . In such cases, Y may be seen as a special case or further evolution of X, which some reviewers do not subscribe to. These two proposals should therefore be put into different initial clusters.

<sup>9</sup> They may however be merged in Phase 3 to form a consolidated cluster but this is a distinctly different process.



### 3.3 Method – Manual Clustering

In some cases, this method may not yield a reliable, unambiguous result. Reasons may be small numbers of available ratings per proposal, or ambiguous ratings. Such proposals can easily be detected automatically; here a manual cluster assignment is indicated. A participant, either from a group of key users or randomly selected altogether, is invited to manually assign the proposal to a cluster. Because requiring much more efforts from the selected participants, this manual interaction should however be limited in number of interventions, the clustering should mainly rely on the automatism outlined in Section 3.2.

### 3.4 Pre-Selection for the Lead Proposal

The lead proposal within a cluster will be determined in Phase 2, however, a preliminary ranking can also be achieved in Phase 1 by ranking the proposals in a cluster according to the (average) quality grades. Allowance must be made for the number of rankings, so that a proposal that only received very few (but high) quality rankings does not supersede a widely-ranked but not so highly placed proposal. Solutions could be a threshold for the number of rankings the proposal must have to become the lead proposal of the cluster or a weight by which widely-ranked proposals are given priority.

## 4. Advantages and Further Research

The method presented herein gives voice to minorities, whose opinions are not supported by many. These opinions will remain visible in a form of distinct clusters, where such clusters will have fewer participants and probably also fewer ratings, nevertheless “survive” the initial Phase 1. It also avoids “claque” effects by inviting reviews and by not basing the chances of survivability of an idea on the number of comments or “likes” or similar things.

Another advantage is the highly automated clustering process and the fact that manual interaction in this regard can be done by participants who are randomly selected. This avoids a “committee of moderators”, who would effectively be in charge of the clustering process and thereby exercise a degree of control that may seem inappropriate to a large portion of the participants, who may prefer a self-organising community to a hierarchy of moderators.

“The proof of the pudding is in the eating”. This also applies to methods of social interaction via digital systems. The exact algorithmic implementation and the fine-tuning of the parameters used is a crucial and necessary empirical validation of the method proposed in this paper. The next step therefore can only be a pilot implementation of the whole system proposed by the authors including the ranking and clustering process proposed in this contribution.

## 5. References

All web quotes checked per March 25, 2017.

- [1] KOUSSOURIS, S., CHARALABIDIS, Y. and ASKOUNIS, D., A Review of the European Union, eParticipation action pilot projects in: PROSSER, A. (ed.), *Transforming Government: People, Process and Policy – Special Issue on Electronic Citizen Participation – state of the art*, Vol. 5(1), 2011.

- 
- [2] EISEL, S.: Internet-basierte Bürgerhaushalte nicht überschätzen; Konrad Adenauer Stiftung, at <http://www.kas.de/wf/de/71.9809/>
- [3] FISHKIN, J.S.: *Democracy and Deliberation: New Directions for Democratic Reform*. Yale University Press, New Haven, CT, 1991.
- [4] FISHKIN, J.S.: *The Voice of the People: Public Opinion and Democracy*. Yale University Press, New Haven, CT.
- [5] VELIKANOV, C. and PROSSER, A.: Mass Online Deliberation in Participatory Policy-Making – Part I in: PAULIN, A.: *Beyond Bureaucracy*, Springer, Berlin, 2017, forthcoming.
- [6] VELIKANOV, C.: Deliberativeness and other important characteristics of e-participation. Proceedings of the EDEM Conference 2011, Austrian Computer Society, Vienna, 2011.
- [7] FILLMORE-PATRICK, H.: *The Iceland Experiment (2009–2013): A Participatory Approach to Constitutional Reform*. Democratization Policy Council (DPC) Policy Note, New Series # 02, Sarajevo, 2013.
- [8] LANDEMORE, H.: *Inclusive Constitution-Making: Epistemic Considerations on the Icelandic Experiment*. European Consortium for Political Research Convention, Bordeaux, France, September 2013, download from <https://press.ecpr.eu/Events/PaperDetails.aspx?PaperID=5097&EventID=5>.
- [9] JONSSON, M.E.: *Democratic Innovations in Deliberative Systems – The Case of the Estonian Citizens’ Assembly*. *Journal of Public Deliberation*, Vol. 11 (1), 2015.
- [10] AITAMURTO, T. and LANDEMORE, H.: *Five design principles for crowdsourced policymaking: Assessing the case of crowdsourced off-road traffic law in Finland*. *Journal of Social Media for Organizations*, Vol. 2(1), 2015.
- [11] VELIKANOV, C. and PROSSER, A.: *Mass Online Deliberation in Participatory Policy-Making – Part II* in: PAULIN, A.: *Beyond Bureaucracy*, Springer, Berlin, 2017, forthcoming.
- [12] *National Socialist Party of America et al. v. Village of Skokie*, U.S. Supreme Court Docket No. 76-1786, 97; S. Ct. 2205; 53 L. Ed. 2d 96.
- [13] MÜLLER-TÖRÖK, R.: *Falschmeldungen: Verbotsvorschläge bewegen sich jenseits aller Realität*, in: *Staatsanzeiger*, 27. Januar 2017, Stuttgart.
- [14] PEUKERT, A.: *Kurzer Prozess mit der Meinungsfreiheit; Cicero*, download at <http://cicero.de/berliner-republik/netzwerkdurchsetzungsgesetz-kurzer-prozess-mit-der-meinungsfreiheit>
- [15] *Communications Assistance for Law Enforcement Act (CALEA)*; Pub.L. 103–414, 108 Stat. 4279, H.R. 4922.
- [16] *SÜDDEUTSCHE ZEITUNG: Eine Stadt verschwindet*, online edition 17/2/2012, retrieved at <http://www.sueddeutsche.de/muenchen/facebook-auftritt-von-muenchen-eine-stadt-verschwindet-1>.

- 
- [17] DERSTANDARD: Facebook:StrachenichtwegenNSA-Kritikgesperrt. Online edition 17/2/2012, retrieved at: <http://derstandard.at/1373512300414/Facebook-sperrt-HC-Strache-aus>.
- [18] ERTL, S.: Protest als Ereignis – Zur medialen Inszenierung von Bürgerpartizipation. transcript Verlag, Bielefeld, 2015.
- [19] NOBLE, D. and RITTEL, H.W.J.: Issue-Based Information Systems for Design. Proceedings of the ACADIA '88 Conference, Association for Computer Aided Design in Architecture, University of Michigan.
- [20] Web Accessibility Initiative: WAI Guidelines and Techniques; at <https://www.w3.org/WAI/guid-tech>
- [21] Bundesgesetz über Regelungen zur Erleichterung des elektronischen Verkehrs mit öffentlichen Stellen (E-Government-Gesetz - E-GovG); available at <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20003230>
- [22] SPERONI DI FENIZIO, P. and VELIKANOV, C.: System Generated Requests for Rewriting Proposals. Proceedings of the ePart 2011 conference; download from <https://arxiv.org/abs/1611.10095>