Building Online Public Consultation Knowledge Graphs

William Aboucaya[1,*], Sonia Guehis[2,4] and Rafael Angarita[3,4]

[1]INRIA Paris
[2]LAMSADE – Université Paris Dauphine
[3]LIP6 – Sorbonne Université
[4]Université Paris Nanterre

Abstract

Online consultation platforms have improved the possibilities for citizens to have an input on public decision making. However, and especially at large scale, identification of the topics discussed and entities evoked has been identified as difficult for both citizens and platform administrators. In this paper, we leverage topic modeling, Named Entity Recognition and Linking and Semantic Textual Similarity to build a knowledge graph representing the different contributions to the *République Numérique* online citizen consultation in French language. The generated graph links the different proposals to topics identified in the consultation and to relevant DBpedia resources. The model proposed for representation of citizen consultations as knowledge graphs simplifies the retrieval of proposals focused on specific topics or mentioning a given entity. It also allows us to improve contextualization of important words in proposals by linking them to short definitions extracted from Wikipedia.

Keywords

Knowledge graph, citizen participation, topic modeling, semantic textual similarity, DBpedia, French language

1. Introduction

Citizen participation has become increasingly popular in recent decades [1, 2] to involve citizens in public decision-making [3]. More specifically, in comparison with their offline counterparts, online participatory platforms allow to gather a larger, more diverse audience, therefore helping to produce results more representative of the consulted population's opinions. Online participatory platforms have been used in a wide variety of use cases, such as participatory budgeting [4, 5], citizen consultation [6, 7], or gathering and release of open data produced by citizens [8]. These platforms have addressed a wide range of issues, such as public space planning [9], digital transformation of countries [10], and climate change mitigation and adaptation [11, 12].

TEXT2KG 2023: Second International Workshop on Knowledge Graph Generation from Text, May 28 - Jun 1, 2023, co-located with Extended Semantic Web Conference (ESWC), Hersonissos, Greece
*Corresponding author.
william.aboucaya@inria.fr (W. Aboucaya); sonia.guehis@parisnanterre.fr (S. Guehis); rafael.angarita@lip6.fr (R. Angarita)
https://rafaelangarita.github.io/ (R. Angarita)
0000-0003-2413-6968 (W. Aboucaya); 0000-0002-2025-2489 (R. Angarita)

Consultations are a means of allowing citizens to offer input on a given topic to guide the actions taken by a given institution. Even though institutions are not necessarily constrained by consultation outcomes, consultations are often regarded as an important means of improving citizens' impact on public policies. Online consultation platforms typically allow users to express their ideas using text and sometimes other media. These proposals are then either directly sent to the institution or shared publicly to solicit feedback from citizens using an upvote/downvote system or written comments.

Online citizen participation platforms have been used in various countries [11] and frequently used consultation techniques. Examples of online citizen consultations at the local scale [13], which are the most common, generally focus on urban planning or local policies. An example of a large-scale online consultation can be found in the *République Numérique* (Digital Republic, *RepNum* for short in the following) consultation. This consultation aimed at letting citizens and other stakeholders publish opinions and amendments to a draft bill focused on the digital transformation of the French administration, open data, internet access, data protection, and open science. By the end of the consultation, 692 proposals had been submitted on the platform.

One of the challenges of using public consultations is managing the high volume of information that they generate, especially in large-scale consultations [14, 15]. This information overload can be overwhelming for both citizens and platform administrators. Citizens may have difficulty finding contributions related to their areas of interest or expertise, while platform administrators may struggle to read and summarize all of the contributions. This challenge is an open one that has been studied by researchers and experts in the field. In addition, this issue is exacerbated by the fact that most participatory platforms rely on simple methods for information retrieval, such as text search without relevance assessment or complex criteria. However, using preexisting resources from a knowledge graph has been proven to be an effective tool for information retrieval, as demonstrated by previous research [16, 17]. Therefore, this approach can be used to address the issue of information overload faced by stakeholders in participatory platforms.

Other issues have also been identified such as the difficulty of reaching a consensus in a collaborative environment and the absence of effective communication between institutions and citizens [18].

In comparison with existing search systems in public consultation platforms, the use of a knowledge graph facilitates collaboration between users to improve the quality of debates and can lead to an accurate decision-making support tool. Moreover, graph-based algorithms, recommender systems, and semantic reasoning based on users' votes and comments can enable advanced data analytic and reasoning capabilities to identify the most relevant proposals by topic and cross topics [19, 20].

In this work, we propose to build a knowledge graph of public consultations to address the challenges discussed above and enhance such participatory processes. Its contributions are the following: *i)* the enrichment of online public consultation data with linked data to improve the contextualization of texts, facilitate its integration with other linked data sources including other consultations, increase transparency and accountability, and enhance decision making; *ii)* the analysis of proposals to infer their topics for better organization and understanding of the information contained within the consultations. By categorizing consultations based on topics, it can be easier to identify trends and patterns in public opinion and to compare and contrast different viewpoints; and *iii)* we provide recommendations for future work based on the specific characteristics of the public consultations of our datasets.

The rest of the paper is structured as follows. Section 2 gives an overview of some of the relevant related works. In section 3, we describe the approach adopted to build the online public consultation knowledge graph. Section 4 discusses the proposed solution and dresses perspectives for future work. Code used in the production of this paper is available at https://github.com/WilliamAboucaya/repnum-kg.

2. Related Work

Knowledge Graphs (KGs) represent facts in the form of nodes and relationships [21, 22]. They use a graph-based data model to capture knowledge in scenarios that involve integrating, managing and extracting data from diverse sources [23]. The two main graph models are RDF [24] and Property Graphs [25]. RDF is a W3C standard RDF for data interchange on the Web. Nodes are resources with a unique identifier on the Web. Edges link the nodes by building subject–predicate–object expressions. Property graphs are directed labeled multigraphs in which nodes and edges can have properties in the form of key-value pairs. Nodes represent entities and edges represent relationships between those entities. Both RDF and Property Graphs have their advantages and disadvantages, depending on the use case [26, 27].

The creation of a KG from text involves the use of Natural Language Processing (NLP) techniques, including Named-Entity Recognition (NER) [28] and Named-Entity Linking (NEL) [29]. NER focuses on identifying and categorizing entities within text, such as people, organizations, and locations, among others. NEL, on the other hand, connects the recognized entities to entities within a knowledge base. Several methods for constructing domain-specific KGs have been developed across a range of domains, including fashion, medicine, and more [30, 31, 32, 33].

Barroca *et. al.* [30] applied Named Entity Recognition (NER) and Named Entity Linking (NEL) techniques on product textual descriptions to enhance a fashion-related KG. The process involved identifying entities, such as materials used in the product, within the descriptions and linking them to their corresponding, unique nodes in the KG. Rincon-Yanez *et. al.* [31] proposed a methodology for creating KGs related to novels. This methodology involves using NLP techniques such as NER and NEL to generate simple triples as a starting point, and then enhancing them using external open knowledge sources, such as DBpedia, to build a comprehensive KG. This methodology was tested on a novel, resulting in the extraction of characters and their relationships. In a similar vein, Alpizar-Chacon and Sosnovsky [34] proposed a solution which automatically labels book's index item in relation to the main subject of a textbook area. These index concepts can be associated with the central themes of the book, common concepts, topics related to but distinct from the book's subject area, or unrelated to the book's subject area altogether.

Reklos and Meroño-Peñuel [32] developed an approach for identifying causal relationships in medical publications. Their approach involves detecting causal sentences, extracting entities, and using this information to build a KG that represents the causal relationships. For example, a node representing the cause of an illness would be connected to a node representing the illness.

Another relevant application of KGs is the analysis of parliamentary debates, as demonstrated

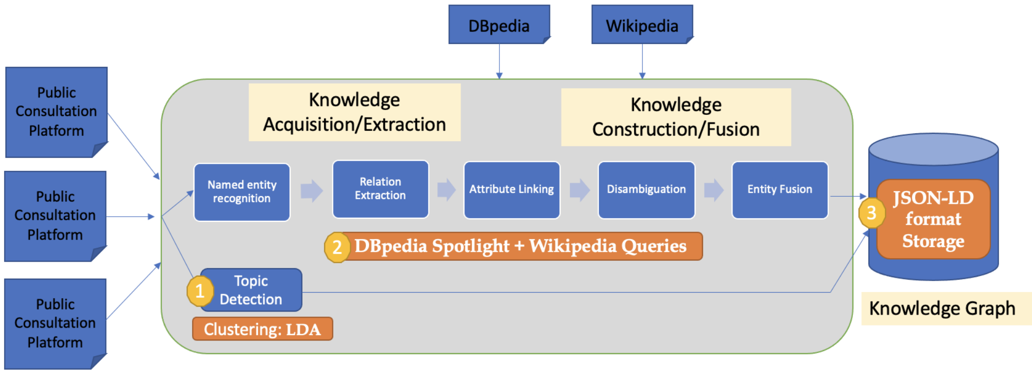

Figure 1: Steps for building the Online Public Consultation KG.

by Tamper *et. al.* [33]. They proposed a method for constructing and enhancing KGs of plenary debates in the Parliament of Finland. The methodology involves preprocessing the text, including lemmatization to normalize the text and subject indexing to describe the text succinctly, due to the particularities of the Finnish language. NER and NEL techniques are then applied to extract named entities and link them to the KG.

NLP techniques have been utilized in various ways to improve public consultations [35, 36, 37]. Weng *et. al.* [35] aimed to aggregate individual contributions into common narratives for improved collaboration by extracting sentences in the form of subject + verb + object and verb + object. Similarly, the New Zealand government categorized proposals in a public consultation using NLP, as per [36]. Another NLP application in a public consultation was done in collaboration with the Italian Ministry of Education, where key concepts and frequently used verbs were extracted from the answers, as demonstrated by Caselli *et. al.* [37].

In [38], Cantador *et. al.*proposed an approach to identify controversial proposals in a consultation by utilizing external knowledge extracted from open government data collections. Moreover, for building the thematic taxonomy, authors manually set categories and assign a platform's pre-existing tags to appropriate categories. Despite the progress made using NLP techniques on public consultations, to the best of our knowledge, our proposed methodology for constructing online public consultation KGs based solely on the textual content submitted by contributors is a novel approach.

3. Building an Online Public Consultation Knowledge Graph

Figure 1 shows the different stages of our methodology for building the consultation KG. We detail these stages in the rest of this section by building a KG for the *RepNum* consultation. This consultation took place from September 26th, 2015 to October 18th, 2015, and received 692 proposals, including the 30 original articles of the bill generated by the organizing ministry. The consultation data is publicly available as open data through the following link: https://www.data.gouv.fr/fr/datasets/consultation-sur-le-projet-de-loi-republique-numerique. It includes the contributions of 21,464 users, almost 150,000 votes, and 6,000 arguments – which are comments

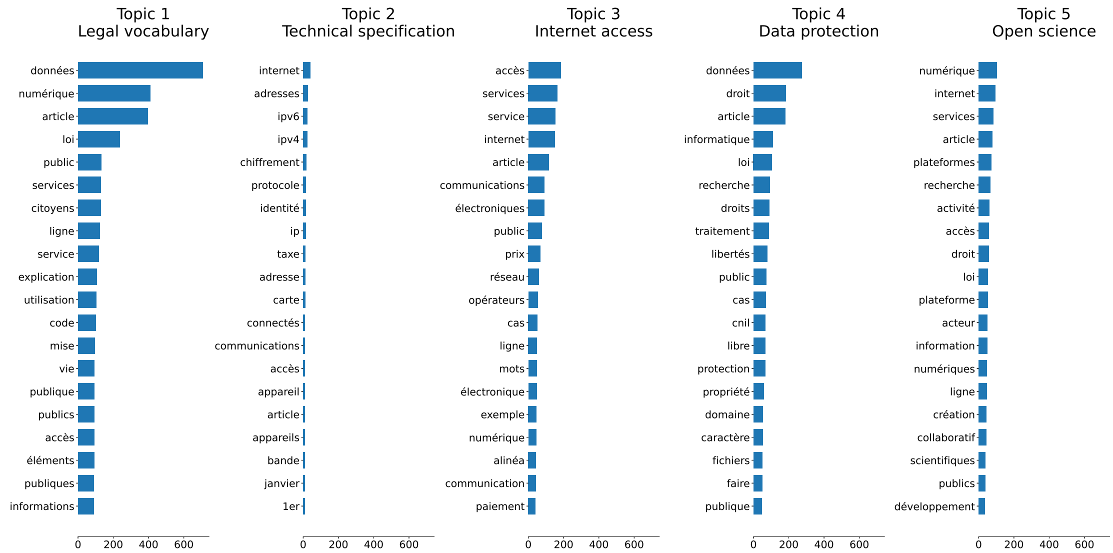

Figure 2: LDA representation of the main topics discussed in the *République Numérique* citizen consultation and their most frequent words. Topic labels were chosen a posteriori by the authors of the paper to help differentiate them.

identified as for or against a proposal.

3.1. Topic Modeling

We use Latent Dirichlet Allocation (LDA) [39] on the proposals submitted to the *RepNum* consultation to identify the different topics. Each proposal is represented as a document, and LDA is an algorithm for topic modeling in a set of documents where each topic is associated with the different N-grams[1] based on their number of occurrences in the different documents. Each proposal is then associated with all relevant topics based on a truth value score. Using this algorithm, we need to choose the number k of topics to identify. We applied LDA for $k \in [4, 15]$ and assessed the suitability of each version of the topic modeling based on the consistency of the 20 most frequent words for each topic. We consider $k = 5$ to be the parameter giving the most coherent results for the *RepNum* citizen consultation. Figure 2 shows the five main topics identified and their most frequent words. We use these topics, in addition to the proposal's title, author, and content, to further characterize the different proposals.

3.2. Entity Recognition and Linking

We use a French-language adaptation of DBpedia Spotlight [40, 41] (DBS for short in the following) to identify entities in our proposals. DBS extracts important words from a given text and links them to resources in DBpedia [42]. We have set the "confidence" parameter to 0.6 and left everything else to default. Our observations about DBS concern the French model. Linking

[1] A sequence of 1 or more words usually written consecutively. For example, both "text" and "knowledge graph" are N-grams in the context of this workshop.

the most important elements in the proposals to DBpedia resources allows us to associate these elements with definitions from Wikipedia and connect proposals based on the resources they mention. To eliminate certain common false associations identified in preliminary tests, we filter the obtained results:

- DBS tends to identify years and dates as important elements of the proposals and link them with the aggregation Wikipedia page for events happening at this given date - e.g., https://en.wikipedia.org/wiki/December_13-. In the context of an online consultation, such elements are often mentioned in contexts where this linking is not relevant. Therefore, we exclude these resources before performing NEL.
- We also observe that the model tends to identify certain N-grams as film titles. This has almost only led to incorrect identification of common N-grams as film titles in the context of the *RepNum* consultation. Consequently, we have filtered non-documentary films out of the resource graph with which we associate our entities. The choice to keep documentary films is made to give citizens the possibility to add references to support their proposals.

Then, we identified an additional issue: acronyms with multiple definitions are often associated with the wrong resource. For example, the acronym "CADA" is used in a proposal to mention the "Administrative Documents Access Commission" (in French, *Commission d'Accès aux Documents Administratifs*). However, this entity is linked to "Shelter for Asylum Seekers" (in French, *Centre d'Accueil de Demandeurs d'Asile*), even though "administrative documents" is mentioned further in the proposal. This issue has been confirmed with similar results by using DBS on another online consultation called RUA (Open data available at https://www.data.gouv.fr/fr/datasets/consultation-vers-un-revenu-universel-dactivite-1/). We solve this issue by applying the following process to each acronym identified as important in a proposal.

- If DBpedia has a node with a name identical to the acronym:
  - If this node is a specific resource (e.g., a Wikipedia page) or redirects to it, we associate the acronym with this resource.
  - Else, if the node is a disambiguation node for multiple resources, we gather the abstract for each resource. Then, using the SBERT [43, 44] model for Semantic Textual Similarity mining, we apply cosine similarity to each abstract with the proposal containing the acronym. Finally, we associate the acronym with the resource whose abstract has the highest cosine similarity with the proposal.
- Else, using the Wikipedia API, we search for the acronym and gather the 5 best results. Then, similarly to the previous dash, we use SBERT's cosine similarity to identify the Wikipedia page whose abstract is the most similar to the proposal. Finally, we link the acronym with the DBpedia resource associated with the identified Wikipedia page.

Table 1
NER and NEL results using DBpedia Spotlight on proposals of the RepNum consultation.

Metric	Value
Number of proposals	692
Raw number of annotations	3,463
Number of different entities	1,015
Number of entities linked with exactly 1 proposal	680
Number of entities linked with 5 or more proposals	82
Number of annotations for entities linked with 5 or more proposals	1,655
Number of proposals without any annotation	112
Number of proposals with 5 or more annotations	253
Median number of annotations per proposal	3

3.3. Resulting Knowledge Graph

The methods NER and NEL mentioned earlier generate a set of proposals where certain N-grams are associated with a DBpedia resource. The obtained results are presented in Table 1. A considerable proportion of the proposals (16.2 %) are not associated with any entity, likely because these proposals are shorter than others. Specifically, while the median number of tokens[2] for proposals with zero annotations is 80, it is 188 for proposals with one or more annotations. Moreover, more than one-third of the proposals (36.6 %) are associated with at least five entities. Also, the number of entities linked to at least one proposal is greater than the number of proposals. However, although the number of entities connected to five or more proposals is significantly lower, almost half (47.8 %) of the annotations link proposals to these entities. We store the generated KG using the JSON-LD format [45]. As JSON-LD is an RDF serialization, querying the graph using SPARQL is possible. Figure 3 depicts a general consultation KG built based on this model.

The model we propose facilitates the issuance of new queries in consultations by both citizens and platform administrators. For instance, one can retrieve all proposals mentioning a specific entity while focusing on a particular topic, such as "all the proposals mentioning the National Commission on Informatics and Liberty (CNIL)" with an LDA-score greater than 0.6 for the "Data Protection" topic. Furthermore, the model allows for summarizing contributions by identifying the topics and entities mentioned in a consultation or linking entities to specific groups of users. By using this method, platform administrators can enhance the transparency of their reports on citizen consultations by adding graph-based data, in addition to responses to a selection of proposals[3], either chosen randomly or based on arbitrary criteria. Lastly, our work can be generalized to most online participatory platforms to connect contributions from various online consultations.

[2]In NLP vocabulary, a token refers to a part of a string. In this study, a token is either a number, a word, or a punctuation mark.
[3]For instance, see https://www.republique-numerique.fr/project/projet-de-loi-numerique/step/reponses

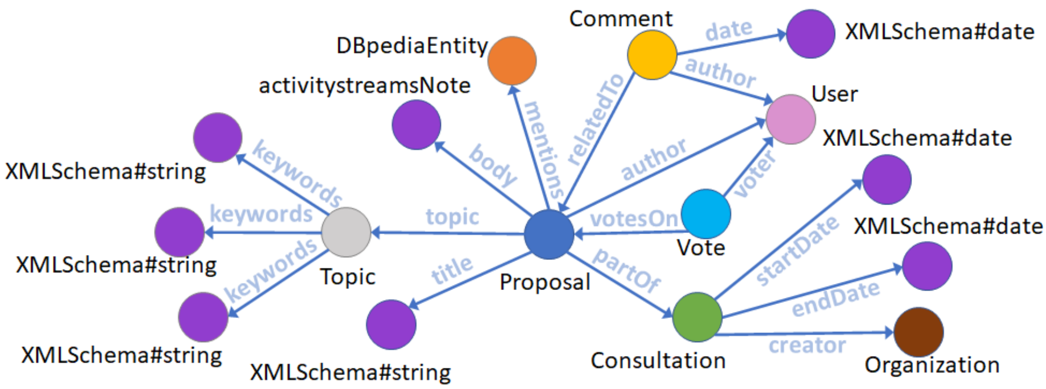

Figure 3: Visualization of a generic consultation stored in JSON-LD.

4. Conclusion and Future Work

Online public consultations allow citizens to create proposals and participate in public debates and decision-making. However, these consultations cover a wide range of topics and are presented in text format, making it challenging for both participatory platform administrators and institutions to manage and extract meaningful insights from them. Additionally, citizens may find it difficult to locate relevant information, identify proposals within a specific topic, or gain an overview of the relationship between existing proposals in a given consultation. To address these challenges, our work proposes a methodology for automatically constructing a consultation knowledge graph that can mitigate these issues. We achieve this by utilizing Latent Dirichlet Allocation (LDA) to identify the most relevant topics in a consultation, and Named Entity Recognition (NER) and Named Entity Linking (NEL) using DBpedia Spotlight to link the different proposals to resources in DBpedia, resulting in a comprehensive and interconnected representation of the consultation data.

In the future, we plan to enhance the consultation knowledge graph by incorporating additional data sources, such as the French National Assembly (https://data.assemblee-nationale.fr), to facilitate the creation of more informed and comprehensive proposals. We also intend to improve our topic modeling methodology by utilizing automated labeling of topics based on Wikipedia articles [46], which could aid in identifying common topics across different consultations. Additionally, we aim to evaluate the ability of DBpedia Spotlight to identify the most important words in proposals. To do so, we plan to perform a quantitative assessment of the relevance of the annotated words and their associated annotations using independent testers.

References

[1] T. Bovaird, Beyond engagement and participation: User and community coproduction of public services, Public Administration Review 67 (2007) 846–860.

[2] A. Lupia, Shortcuts versus encyclopedias: Information and voting behavior in california insurance reform elections, American Political Science Review 88 (1994) 63–76.

[3] H. Baum, Citizen participation, in: N. J. Smelser, P. B. Baltes (Eds.), International Encyclopedia of the Social & Behavioral Sciences, Pergamon, Oxford, 2001, pp. 1840 – 1846.
[4] J. Holston, V. Issarny, C. Parra, Engineering software assemblies for participatory democracy: The participatory budgeting use case, in: 2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C), Institute of Electrical and Electronics Engineers, Austin, TX, USA, 2016, pp. 573–582.
[5] C. Parra, C. Rohaut, M. Maeckelbergh, V. Issarny, J. Holston, Expanding the design space of ict for participatory budgeting, in: Proceedings of the 8th International Conference on Communities and Technologies, C&T '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 213–221.
[6] J. K. Ward, F. Cafiero, R. Fretigny, J. Colgrove, V. Seror, France's citizen consultation on vaccination and the challenges of participatory democracy in health, Social Science & Medicine 220 (2019) 73–80.
[7] W. Aboucaya, R. Angarita, V. Issarny, Detecting Obstacles to Collaboration in an Online Participatory Democracy Platform: A Use-case Driven Analysis, in: Proceedings of the 2nd International Workshop on Equitable Data and Technology, FairWare '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 25–33.
[8] B. Lefevre, V. Issarny, Matching technological & societal innovations: the social design of a mobile collaborative app for urban noise monitoring, in: 2018 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, Institute of Electrical and Electronics Engineers, Taormina, Italy, 2018, pp. 33–40.
[9] A. Wilson, M. Tewdwr-Jones, Let's draw and talk about urban change: Deploying digital technology to encourage citizen participation in urban planning, Environment and Planning B: Urban Analytics and City Science 47 (2020) 1588–1604.
[10] V. Laurent, R. Badouard, C. Mabi, C. Méadel, G. Sire, Les outils numériques gouvernementaux au service de la loi: le design de la plateforme république numérique, Terminal. Technologie de l'information, culture & société (2018).
[11] V. Pina, L. Torres, S. Royo, Comparing online with offline citizen engagement for climate change: Findings from austria, germany and spain, Government Information Quarterly 34 (2017) 26–36. Open Innovation in the Public Sector.
[12] E. Glaas, M. Hjerpe, M. Karlson, T.-S. Neset, Visualization for citizen participation: User perceptions on a mainstreamed online participatory tool and its usefulness for climate change planning, Sustainability 12 (2020).
[13] S. Royo, A. Yetano, B. Acerete, E-participation and climate change: Are local governments actively promoting responsible behaviors and offering opportunities for citizen involvement?, in: 2012 45th Hawaii International Conference on System Sciences, 2012.
[14] K. Chen, T. Aitamurto, Barriers for crowd's impact in crowdsourced policymaking: Civic data overload and filter hierarchy, International Public Management Journal 22 (2019).
[15] O. Perez, Complexity, information overload and online deliberation, Journal of Law and Policy for the Information Society (2009) 10–09.
[16] R. Reinanda, E. Meij, M. de Rijke, Knowledge graphs: An information retrieval perspective, Foundations and Trends® in Information Retrieval 14 (2020) 289–444.
[17] M. Gaur, K. Gunaratna, V. Srinivasan, H. Jin, Iseeq: Information seeking question generation using dynamic meta-information retrieval and knowledge graphs, Proceedings of the

AAAI Conference on Artificial Intelligence 36 (2022) 10672–10680.
[18] M. Ianniello, S. Iacuzzi, P. Fedele, L. Brusati, Obstacles and solutions on the ladder of citizen participation: a systematic review, Public Management Review 21 (2019) 21–46.
[19] M. Y. Kpiebaareh, W.-P. Wu, B. Agyemang, C. R. Haruna, T. Lawrence, A generic graph-based method for flexible aspect-opinion analysis of complex product customer feedback, Information 13 (2022).
[20] Q. Guo, F. Zhuang, C. Qin, H. Zhu, X. Xie, H. Xiong, Q. He, A survey on knowledge graph-based recommender systems, IEEE Transactions on Knowledge and Data Engineering 34 (2022) 3549–3568.
[21] M. Lissandrini, D. Mottin, K. Hose, T. B. Pedersen, Knowledge graph exploration systems: are we lost, in: Proceedings of the 12th Conference on Innovative Data Systems Research (CIDR), volume 22, 2022, pp. 10–13.
[22] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. d. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (CSUR) 54 (2021) 1–37.
[23] N. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, J. Taylor, Industry-scale knowledge graphs: Lessons and challenges: Five diverse technology companies show how it's done, Queue 17 (2019) 48–75.
[24] F. Manola, E. Miller, B. McBride, et al., Rdf primer, W3C recommendation 10 (2004) 6.
[25] R. Angles, The property graph database model., in: AMW, 2018.
[26] D. Alocci, J. Mariethoz, O. Horlacher, J. T. Bolleman, M. P. Campbell, F. Lisacek, Property Graph vs RDF Triple Store: A Comparison on Glycan Substructure Search, PLOS ONE 10 (2015) e0144578.
[27] A. Donkers, D. Yang, N. Baken, Linked data for smart homes: Comparing rdf and labeled property graphs, in: M. Poveda-Villalón, A. Roxin, K. McGlinn, P. Pauwels (Eds.), LDAC 2020 Linked Data in Architecture and Construction, CEUR Workshop Proceedings, CEUR-WS.org, 2020, pp. 23–36. 8th Linked Data in Architecture and Construction Workshop, LDAC 2020, LDAC 2020 ; Conference date: 17-06-2020 Through 19-06-2020.
[28] D. Nadeau, S. Sekine, A survey of named entity recognition and classification, Lingvisticae Investigationes 30 (2007) 3–26.
[29] W. Shen, J. Wang, J. Han, Entity linking with a knowledge base: Issues, techniques, and solutions, IEEE Transactions on Knowledge and Data Engineering 27 (2014) 443–460.
[30] J. Barroca, A. Shivkumar, B. Q. Ferreira, E. Sherkhonov, J. Faria, Enriching a fashion knowledge graph from product textual descriptions, in: Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text co-located with 19th Extended Semantic Conference (ESWC 2022), Hersonissos, Greece, May 30th, 2022, volume 3184 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 36–51.
[31] D. Rincon-Yanez, S. Senatore, FAIR knowledge graph construction from text, an approach applied to fictional novels, in: Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text co-located with 19th Extended Semantic Conference (ESWC 2022), Hersonissos, Greece, May 30th, 2022, volume 3184 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 94–108.
[32] I. Reklos, A. Meroño-Peñuela, Medicause: Causal relation modelling and extraction from medical publications, in: Proceedings of the 1st International Workshop on Knowledge

Graph Generation From Text co-located with 19th Extended Semantic Conference (ESWC 2022), Hersonissos, Greece, May 30th, 2022, volume 3184 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 1–18.

[33] M. Tamper, R. Leal, L. Sinikallio, P. Leskinen, J. Tuominen, E. Hyvönen, Extracting knowledge from parliamentary debates for studying political culture and language, in: Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text co-located with 19th Extended Semantic Conference (ESWC 2022), Hersonissos, Greece, May 30th, 2022, volume 3184 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 70–79.

[34] I. Alpizar-Chacon, S. Sosnovsky, What's in an index: Extracting domain-specific knowledge graphs from textbooks, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 966–976.

[35] M.-H. Weng, S. Wu, M. Dyer, Ai augmented approach to identify shared ideas from large format public consultation, Sustainability 13 (2021) 9310.

[36] N. Stats, Ethnicity standard classification: Findings from public consultation november 2019, Stats NZ: Wellington, New Zealand (2020).

[37] T. Caselli, G. Moretti, R. Sprugnoli, S. Tonelli, D. Lanfrey, D. S. Kutzmann, Nlp and public engagement: The case of the italian school reform, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 401–406.

[38] I. Cantador, M. E. Cortés-Cediel, M. Fernández, Exploiting open data to analyze discussion and controversy in online citizen participation, Information Processing & Management 57 (2020) 102301.

[39] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, L. Zhao, Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey, Multimedia Tools and Applications 78 (2019) 15169–15211.

[40] P. N. Mendes, M. Jakob, A. García-Silva, C. Bizer, Dbpedia spotlight: Shedding light on the web of documents, in: Proceedings of the 7th International Conference on Semantic Systems, I-Semantics '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 1–8.

[41] J. Daiber, M. Jakob, C. Hokamp, P. N. Mendes, Improving efficiency and accuracy in multilingual entity extraction, in: Proceedings of the 9th International Conference on Semantic Systems, I-SEMANTICS '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 121–124.

[42] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, Dbpedia: A nucleus for a web of open data, in: K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, P. Cudré-Mauroux (Eds.), The Semantic Web, Springer Berlin Heidelberg, 2007, pp. 722–735.

[43] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using Siamese BERT-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992.

[44] N. Reimers, I. Gurevych, Making monolingual sentence embeddings multilingual using

knowledge distillation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2020.
[45] M. Sporny, D. Longley, G. Kellogg, M. Lanthaler, N. Lindström, Json-ld 1.1, W3C Recommendation, Jul (2020).
[46] J. H. Lau, K. Grieser, D. Newman, T. Baldwin, Automatic labelling of topic models, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Portland, Oregon, USA, 2011, pp. 1536–1545.