Visual exploration of digital cultural artifacts (short paper)

Extended Abstracts of doctoral thesis

Eleonora Bernasconi[1], Miguel Ceriani[2] and Massimo Mecella[1]

[1]*Sapienza Università di Roma, Dipartimento di Ingegneria Informatica, Automatica e Gestionale Antonio Ruberti, via Ariosto, 25, 00185 Rome, Italy*

[2]*Università degli Studi di Bari Aldo Moro, Dipartimento di Informatica, via Edoardo Orabona, 4, 70125 Bari, Italy*

Abstract

This work presents the extended abstract of the doctoral thesis about the Visual Exploration of digital cultural artifacts. We exploited the synergy between semantic technologies and knowledge exploration tools for digital libraries, proposing a system which includes: semantic enrichment of textual content; a user interface for search and exploration of a digital library through visual navigation of a knowledge graph of topics; the extensions of the central system to leverage the extracted knowledge. For example, interface components hide complex queries to find interesting topics for users; an integrated system to semi-automatically manage the knowledge contained in the images, generating video trailers of books; and a collaborative system to improve extracted data quality.

Keywords

Semantic Web, Knowledge Graph, Visual Search Interface, Digital Library, Digital Humanities,

1. Research Context

Searching and exploring a vast text corpus has often arisen as a human need. Traditionally, the search process is based on manually curated metadata classifying documents by arguments, authors, metadata, etc. Albeit the metadata that used to be stored in physical cabinets is now stored in databases, the process often remains similar.

Although being a decisive paradigm, the maintenance of metadata is costly and becomes progressively more expensive and less reliable with the increase of required detail. The transition to electronic documents (either created natively as such or digitized) enables the direct text-based search of the content. Text-based search for the full content of documents is a powerful tool. However, it comes with its limitations due to the inherent ambiguity of natural languages and the need for the user to anticipate the actual words used in the content, as the machine cannot capture what the user and the corpus *mean*. This is called the semantic gap. Statistical methods can be successfully used for query expansion, mitigating the issue, but the user has

1st Italian Workshop on Artificial Intelligence for Cultural Heritage (AI4CH22), co-located with the 21st International Conference of the Italian Association for Artificial Intelligence (AIxIA 2022). 28 November 2022, Udine, Italy.

bernasconi@diag.uniroma1.it (E. Bernasconi); miguel.ceriani@uniba.it (M. Ceriani); mecella@diag.uniroma1.it (M. Mecella)

0000-0003-3142-3084 (E. Bernasconi); 0000-0002-5074-2112 (M. Ceriani); 0000-0002-9730-8882 (M. Mecella)

no control of the process. Semantic enrichment methods, as *named-entity recognition and linking (NERL)* [39, 40], aim to bridging the semantic gap between raw text and concepts, by associating words in the documents with entities in a knowledge base, often a knowledge graph (KG). NERL successfully enabled users to search and analyze text corpora [41] more effectively. Nevertheless, the navigation of semantic relationships (with their meaning, rather than just as generic connections) between extracted entities has seldom been adopted as a method for the exploration of a corpus; even if it is known that the cognitive processes in library searching are generally more complicated than a single topic-based search [42]. Also, while knowledge extraction methods as NERL are now broadly used by big players in the industry as well as in academic projects, their usage by small to medium size organizations (which often have text corpora, either private or public, that they struggle to manage in a structured way consistently) is still minimal, in part due to the lack of an established standard workflow. The idea proposed in this thesis to address the problems identified concerns a system that, through the use of:

- artificial intelligence techniques, which extract information from unstructured sources (such as text and images);
- semantic technologies that give an unambiguous meaning to the extracted concepts and connect the information with other knowledge bases;
- an interface for explore and query the knowledge graph;
- an interaction paradigm which supports the serendipity effect to discover unexpected things.

allows anyone possessing a digital library to extrapolate, reuse and promote its contents.

Building and maintaining institutional systems of organization of knowledge and related datasets in libraries required years of work for a highly skilled and trained workforce. Knowledge graphs in the domain of libraries and digital humanities demonstrate how the application of automatic knowledge extraction and semantic enrichment to large-scale corpora opens up a spectrum of possible new research questions that, until now, were difficult to answer with existing methods.

Exploiting the opportunities in the digital humanities research field poses many methodological and technical challenges.

- Novel user interface and interaction paradigm are needed to support users in viewing, annotating, and systematically analyzing relevant parts of possibly large digitized corpora. Users could express relevance by selecting corresponding concept definitions in knowledge graphs.
- Scalable text-mining and machine-learning techniques are needed to systematically and efficiently analyze and compare the characteristics, contents, and relationships of concepts expressed in knowledge graphs within and across corpora.
- Algorithms are needed that support users in detecting, contextualizing, and analyzing various forms of expressions and associated narrative techniques in corpora spanning an extended period, in which the syntax and semantics may have been subject to constant change.

For these reasons, the need arises to propose:

- the development of tools and scalable techniques for aligning large-scale, multi-media corpora with concepts expressed in knowledge graph;
- support in knowledge exploration through a novel interaction paradigm based on the principle of serendipity that enables users to discover unexpected things.
- validation mechanisms ensure trust in data quality when humans curate data with different levels of expertise and result from automatic processes.

2. Research Objectives and contributions

The various research phases related to the study of the state of art, the identification of open challenges and the design development and evaluation process of the proposed solution will be discussed in detail below.

2.1. Studying the state-of-the-art

A digital library's knowledge extraction and management are receiving increasing attention in industrial and academic research fields. Thousands of publications use artificial intelligence and semantic technologies tools to reach the research goals.

Forty-four tools were selected and classified by: interaction paradigm used, such as Node-Link visualization [24, 27, 28, 23], tabular visualization or visual query composition [34, 16, 20]; type of information viewed (data [31], model visualization [18, 17], data to model visualization [32, 19, 33]; complexity reduction strategies [29, 26, 30, 22] to optimize the search results viewed; background characteristics, such as semantic enrichment and knowledge extraction [53, 51, 48, 55, 52]; the use of semantic annotation (automatic [47], semi-automatic [50, 54] and manual [49]). Furthermore, many tools face the challenge of exploring the contents of a digital library, but two, in particular, go in the same direction as this work. Yewno Discover [45] is an integrated system that offers classification and visual exploration of academic materials to help scholars in their research. However, it is not adaptable and flexible to different contexts of use, except with ad hoc adjustments. Furthermore, concerning the proposed system, it makes limited use of the KG structure for exploration, which is at the core of the research questions posed here.

Sampo-UI [46] is a framework that provides a set of reusable and extensible components, application state management, and a read-only API for SPARQL queries, which can be used to create a user interface for a semantic portal. Unlike Sampo UI, the system proposed in this work also offers a knowledge extractor service from unstructured data and a semantic enrichment service.

From literature emerges like the knowledge graphs are a prominent answer to disseminating cultural heritage challenge. The nature of KG is integrability[38]. This feature allows connecting different cultural domains on the web in the form of linked open data, thus promoting the dissemination of cultural heritage. Furthermore, KG can be explored and interrogated with complex queries favouring the discovery of new knowledge (serendipity).

Compared to the tools analyzed so far, we propose a tool with the incremental exploration of knowledge graph information (supporting the principle of serendipity) with the novelty of application to the exploration context of a digital library. Our system aims to facilitate the

search, viewing and exploration of books or documents that deal with the information sought. Among the tools analyzed, no system focuses on exploring a catalogue of books that exploits the interaction paradigm proposed in this work and detailed in section 2.2.

2.2. Designing and developing an original system

Once that concrete open challenges have been individuated, the pursued task is solving them. The good methodological practices learned in the research path bring ideas and instruments to approach such problems. However, the main contribution of this research was to go beyond.

From the literature and a study conducted with five researchers belonging to humanities, common behaviours were identified for searching a digital library for content[11, 37, 12].

The main general supported search behaviors are the following:

- find documents relevant to a specific topic;
- expand or specialize searches by moving through related topics;
- have visibility of available related resources, which could potentially be of interest;
- visually organize the resources found by considering their relationships and properties;
- find topics and documents at the crossing of multiple topics, possibly of different kinds (places, people, time periods, etc.).

For the sake of the analytic approach, the experimentation effort was framed through a set of research goals[10].

- Would users exploring a corpus of text profit from the semantic navigation of the associated KG of topics?
- What kind of user interface would effectively support such a navigation?
- What kind of users, scenarios, and tasks would benefit from this interaction paradigm?
- Does building and maintaining a semantic enrichment and KG creation pipeline necessarily involve high upfront costs and highly skilled developers?

To answer the above questions, the following hypotheses have been formulated.

- Users will be able to effectively explore a text corpus through a KG-based user interface, which offers the following main functions:
  - *a.* finding concepts through text search (among the ones pertinent to the specific domain),
  - *b.* visually navigating the concepts and their relationships, and
  - *c.* showing documents relevant to the selected concept.
- The method, given a corpus of texts in a specific domain, will benefit both users with little knowledge of the domain (by supporting semantically-relevant discovery) and domain experts (by enabling a topic-oriented visual organization of the documents).
- It is feasible to build a ready-to-use complete system, including both semantic enrichment pipeline and web-based front end, which is able, with only some configuration, to be applied to any specific corpus to enable the KG-based exploration.

While the first two research questions and related hypotheses are relevant for investigating the benefits of the proposed approach for the end users, the last research question and hypothesis investigate the usefulness and portability of such a system to different contexts of use.

2.2.1. Evaluation of the proposed system

We evaluated the proposed solution, based on a corpus search and exploration paradigm with the transparent use of knowledge graphs to improve and measure the level of acceptance (strengths and weaknesses) by the users of the system. The system has been tested in the context of a specific use case: exploration of the book catalog of medium-size publishing house, specialized in classical antiquity. The anticipated final users of the tool can be roughly classified in two categories:

- domain experts who may adopt a new approach to search and discover resources in the context of their research;
- curious people who want to explore new topics.

The evaluation process[9] lasted two years and was characterised by three phases:

- an evaluation of the extracted data, from the point of view of quality and usefulness, with the help of domain experts;
- a small-scale qualitative user-based evaluation of the tool with a some researchers of the field;
- a larger and richer user-based evaluation of the tool, both on its own and in comparison with other existing solution, which involved both students and researchers of the field.

2.3. Extensions of the proposed system

After the proposed system was positively evaluated, it was decided to proceed by developing some applications of potential interest for the domain of Digital Humanities and Digital Libraries as extensions of the primary system. In particular, the same interaction paradigm was applied to two different domains concerning the humanities (ancient symbols and ancient places). New functionality has been added downstream of the system that processes the images present in the digital library, recognizing the objects represented, which are inserted into the explorable knowledge domain. Also presented is the extension that allows domain experts to validate the quality of automatically extracted reports.

2.3.1. Application of the interaction paradigm to different domain

The availability of a tool such as the one proposed in this work would foster collaboration among the researchers in the area, and could attract curious [36], casual, users by easing the diffusion of niche topics like those regarding ancient documentary texts[7]. Offering a pipeline to build a custom KG, can *(i)* introduce a common vocabulary for researchers in the area, *(ii)* share a common understanding of how concepts are related, *(iii)* enable the reuse of domain knowledge, and *(iv)* make domain assumptions explicit. In addition the graphical user interface can be exploited to allow researchers *(i)* to explore the KG, *(ii)* to search and explore relations and connections between resources, *(iii)* to make historical-geographical implications, and *(iv)* to discover new facts about the research field.

Another extension arises from the researchers' need to explore knowledge bases in a cartographic context[8]. This is done by searching for links between the different topics related to a

place or a key term, in such a way as to reveal unexpected connections during the exploration of contents and, thus, generating new ideas.

2.3.2. Knowledge extraction from images

Another extension of the primary system involved the addition of the automatic extraction of the contents of the images present in a digital library[4, 5, 6]. This extension has allowed the reuse of this information to generate new applications, for example, the semi-automatic creation of book trailers to support storytelling for digital libraries.

Multimedia storytelling is an effective and engaging method to convey information in multiple domains. Specifically, book trailers –video advertisements for books– positively influence the desire to learn and the motivation to read. The video trailer generator system supports an expert by gathering relevant crowd-sourced multimedia content, which, arranged as stories, can be used to showcase a book in the form of video clips. Crucially, the expert controls how the content is finally combined and edited rather than offering a fully automated process.

2.3.3. Validation of the automatic knowledge extraction

After a critical analysis of the entire information extraction and modeling process, the most significant difficulties arise from problems with the quality of automatically extracted data. The limitations found, such as OCR errors and disambiguation errors of the extracted concepts, are part of those problems that limit the system potential, which was attested by the users during the evaluation.

Consequently to the findings, a strategy for improvement was evaluated founded on the following:

- improving improve as much as possible the OCR, NER, and NEL algorithms [15, 21, 25, 35]
- inserting a human control (human in the loop) to validate the automatic extractions.

At least at the moment, no algorithm can extrapolate information with optimal quality, so the human expert needs to have the last word.

For this reason, a proposed solution[3] inserts a layer of human control so that domain experts can validate the automatically extracted results.

Allowing domain experts to collaboratively validate information previously automatically extracted from a Digital Library (DL) is an approach to support the incremental data quality improvement that can be done specifically through the validation of entity linking. Furthermore, rather than seeing just the results of the extraction process, it can be helpful for the domain experts to trace the origin of where the AI recognized a specific entity (i.e. a “snippet” of text or an image).

Acknowledgments

SCIBA (Regione Lazio and MIUR - Determinazione n. G07413 del 16/06/2021) project partly supports this work. Miguel Ceriani acknowledges funding from the European program PON Ricerca e Innovazione 2014-2020, within the project no. COD.AIM1852414, activity 2, line 2.1 and the project “Computational methods for the web economy”, DM 1062/2021.

References

[1] Drucker, Johanna (2021). Digital humanities overview. The Digital Humanities Coursebook, 1-18, Routledge, https://doi.org/10.4324/9781003106531-1.

[2] Gooding, Paul (2020). The Library in Digital Humanities. Routledge International Handbook of Research Methods in Digital Humanities, 137-152, Routledge, https://-doi.org/10.4324/9780429777028-11.

[3] Bernasconi, E., Ceriani, M., Mecella, M., Morvillo, A. Automatic Knowledge Extraction from a Digital Library and Collaborative Validation. International Conference on Theory and Practice of Digital Libraries Springer, Cham. pp. 480–484 Padua (September 2022).

[4] Bernasconi, E., Ceriani, M., Mecella, M., De Luzi, F., Sapio, F. StoryBook. Automatic generation of book trailers. International Conference on Advanced Visual Interfaces Association for Computing Machinery. pp. 1-3 Frascati (June 2022).

[5] Bernasconi, E., Ceriani, M., De Luzi, F., Sapio, F., Mecella, M. Storybook: a tool for the semi-automatic creation of book trailers.Workshop. International Conference on Research Challenges in Information Science. Ceur- ws. Barcelona (May 2022).

[6] Bernasconi, E., Ceriani, M., De Luzi, F., Di Fazio, C., Marcucci, R., Mecella, M., Sapio, F. StoryBook-A Storytelling-based Platform for Digital Book Stores.Workshop. International Conference on Research Challenges in Information Science. Ceur-ws. Barcelona (May 2022).

[7] Bernasconi, E., Boccuzzi, M., Catarci, T., Ceriani, M., Ghignoli, A., Leotta, F., Ziran, Z. Exploring the Historical Context of Graphic Symbols: the NOTAE Knowledge Graph and its Visual Interface. 17th Italian Research Conference on Digital Libraries Ceur-ws. pp. 147-154 Padua (February 2021).

[8] Bernasconi, E., Boccuccia, P., Fabbri, M., Francescangeli, A., Mar- cucci, R., Mecella, M., Morvillo, A., Tondi, E. SCIBA-A Prototype of the Computerized Cartographic System of an Archaeological Bibliography.Workshop. International Conference on Research Challenges in Information Science. Ceur-ws. Barcelona (May 2022).

[9] Bernasconi, E., Ceriani, M., Mecella, M., Catarci, T. Design, realization and user evaluation of the ARCA system for exploring a digital library.Journal. International Journal on Digital Libraries Springer. (2022).

[10] Bernasconi, E., Ceriani, M., Mecella, M. Exploring a Text Corpus via a Knowledge Graph. 17th Italian Research Conference on Digital Libraries Ceur-ws. pp. 91-102 Padua (February 2021).

[11] Ceriani, M., Bernasconi, E., Mecella, M. A streamlined pipeline to enable the semantic exploration of a bookstore. Italian Research Conference on Digital Libraries Springer, Cham. pp. 75-81 Bari (January 2020).

[12] Bernasconi, E., Ceriani, M., Mecella, M. Academic Research Creativity Archive (ARCA). International Conference on Research Challenges in Information Science. Springer, Cham. pp. 713-714 (May 2021).

[13] A. Ghignoli, The notae project: a research between est and west, late antiquity and early middle ages, Comparative Oriental Manuscript Studies Bullettin 5/1 (2019) 27–39.doi:https://doi.org/10.25592/uhhfdm.185.

[14] Bernasconi, E., Ceriani, M., Mecella, M., Catarci, T., Capanna, C., Di Fazio C., Marcucci,

R., Pender, E., Petriccione, F.: ARCA. Semantic exploration of a bookstore. (AVI '20). Association for Computing Machinery, New York, NY, USA, Article 78, pp. 1–3. (2020).
[15] Byrne, Kate (2007). Nested Named Entity Recognition in Historical Archive Text. International Conference on Semantic Computing (ICSC 2007), IEEE, https://-doi.org/10.1109/icsc.2007.107.
[16] F. Haag, S. Lohmann, S. Siek, T. Ertl, QueryVOWL: Visual composition of SPARQL queries, in: Proceedings of ESWC 2015 Satellite Events, in: LNCS, vol. 9341, Springer, 2015, pp. 62–66.
[17] S. Lohmann, V. Link, E. Marbach, S. Negru, WebVOWL: Web-based Visualization of Ontologies, in: EKAW, 2014
[18] Mouromtsev, D., Pavlov, D., Emelyanov, Y., Morozov, A., Razdyakonov, D., Galkin, M.: The simple web-based tool for visualization and sharing of semantic data and ontologies. In: ISWC 2015
[19] M. Weise, S. Lohmann, F. Haag, LD-VOWL: Extracting and visualizing schema information for linked data endpoints, in: Proceedings of the 2nd International Workshop on Visualization and Interaction for Ontologies and Linked Data, VOILA 2016, in: CEUR-WS, vol. 1704, CEUR-WS.org,2016, pp. 120–127.
[20] F. Haag, S. Lohmann, S. Siek, T. Ertl, QueryVOWL: A visual query notation for linked data, in: Proceedings of ESWC 2015 Satellite Events, in: LNCS, vol. 9341, Springer, 2015, pp. 387–402.
[21] Röder, M., Usbeck, R., Ngomo, A. N., (2018). GERBIL – Benchmarking Named Entity Recognition and Linking consistently. Semantic Web, 9(5), 605-625, ISSN 2210-4968, IOS Press, https://doi.org/10.3233/sw-170286.
[22] R. Chawuthai, H. Takeda, RDF Graph Visualization by Interpreting Linked Data as Knowledge, in: JIST, 2015.
[23] F. Viola, L. Roffia, F. Antoniazzi, A. D'Elia, C. Aguzzi, T. Salmon Cinotti, Interactive 3D exploration of RDF graphs through semantic planes, Future Internet 10 (8) (2018).
[24] A.G. Nuzzolese, V. Presutti, A. Gangemi, A. Musetti, P. Ciancarini, Aemoo:exploring knowledge on the web, in: WebSci, 2013.
[25] Chaudhuri, Arindam, Mandaviya, Krupa, Badelia, Pratixa, Ghosh, Soumya K. (2016). Optical Character Recognition Systems. Optical Character Recognition Systems for Different Languages with Soft Computing, 9-41, ISSN 1434-9922, Springer International Publishing, https://doi.org/10.1007/978-3-319-50252-6_2.
[26] D.V. Camarda, S. Mazzini, A. Antonuccio, Lodlive, exploring the web of data, in: V. Presutti, H.S. Pinto (Eds.), I-SEMANTICS 2012 - 8th International Conference on Semantic Systems, I-SEMANTICS '12, Graz,Austria, September 5–7, 2012, ACM, 2012, pp. 197–200.
[27] A. Micsik, Z. Tóth, S. Turbucz, LODmilla: Shared visualization of linked open data, in: Ł. Bolikowski, V. Casarosa, P. Goodale, N. Houssos, P.Manghi (Eds.), Theory and Practice of Digital Libraries – TPDL 2013 Selected Workshops, Springer International Publishing, Heidelberg, 2014, pp. 89–100.
[28] A. Micsik, S. Turbucz, A. Györök, LODmilla: a linked data browser for all, in: S. Harald, F. Agata, L. Jens, H. Sebastian (Eds.),Posters Demos@SEMANTiCS 2014, CEUR-WS.org, 2014, pp. 31–34.
[29] Tuukka Hastrup, Richard Cyganiak, and Uldis Bojars. Browsing linked data with fenfire.

In International World Wide Web Conference (WWW), 2008. 57, 58
[30] L. Po, High-level visualization over big linked data, in: M. van Erp, M.Atre, V. López, K. Srinivas, C. Fortuna (Eds.), Proceedings of the ISWC 2018 Posters Demonstrations, Industry and Blue Sky Ideas Tracks Co-Located with 17th International Semantic Web Conference, ISWC 2018, Monterey,USA, October 8th - to - 12th, 2018, in: CEUR Workshop Proceedings, vol.2180, CEUR-WS.org, 2018.
[31] Marie, N., Gandon, F., Ribiere, M., Rodio, F.: Discovery hub: on-the-fly linked data exploratory search. In: Proceedings of the 9th Int. Conf. on Semantic Systems. pp.17–24. ACM (2013)
[32] C. Anutariya and R. Dangol, "VizLOD: Schema Extraction And Visualization Of Linked Open Data," 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2018, pp. 1-6, doi: 10.1109/JCSSE.2018.8457325.
[33] Jesse C. J. van Dam, Jasper J. Koehorst, Peter J. Schaap, Vitor Martins dos Santos, and María Suárez-Diez. RDF2Graph a tool to recover, understand and validate the ontology of an RDF resource. Journal of Biomedical Semantics, 6:39, 2015. DOI: 10.1186/s13326-015-0038-9 58,63
[34] Florian Haag, Steffen Lohmann, and Thomas Ertl. SparqlFilterFlow: SPARQL query composition for everyone. In Extended Semantic Web Conference (ESWC), pp. 362–367, 2014c. DOI:10.1007/978-3-319-11955-7_49 58, 64
[35] Huynh, Vinh-Nam, Hamdi, Ahmed, Doucet, Antoine (2020). When to Use OCR Post-correction for Named Entity Recognition?. Digital Libraries at Times of Massive Societal Transition, 33-42, ISSN 0302-9743, Springer International Publishing, https://-doi.org/10.1007/978-3-030-64452-9_3.
[36] Mecella, M., Leotta, F., Marrella, A., Palucci, F., Seri, C., Catarci, T.: Encouraging persons to visit cultural sites through mini-games. EAI Endorsed Trans. Serious Games 4(14), e3 (2018).
[37] Bernasconi, E., Ceriani, M., Mecella, M., Catarci, T., Capanna, M.C., Fazio, C.D., Marcucci, R., Pender, E., Petriccione, F.M.: ARCA. Semantic exploration of a bookstore. Proceedings of the International Conference on Advanced Visual Interfaces. (2020).
[38] Hogan, A., et al.Knowledge Graphs. CoRR. https://arxiv.org/abs/2003.02320 (2020).
[39] Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes 30 (1),3–26 (2007).
[40] Shen, W., Wang, J., Han, J.: Entity linking with a knowledge base: Issues, techniques, and solutions. IEEE Transactions on Knowledge and Data Engineering 27(2), 443–460 (2014).
[41] Ristoski, P., Paulheim, H.: Semantic web in data mining and knowledge discovery: A comprehensive survey. Journal of Web Semantics 36, 1–22 (2016).
[42] Dee Andy Michel. 1994. What is used during cognitive processing in information retrieval and library searching? eleven sources of search information. J. Am. Soc. Inf. Sci. 45, 7 (Aug. 1994), 498–514.
[43] Bodenhamer, David J., John Corrigan and Trevor M. Harris. "The Spatial Humanities: GIS and the Future of Humanities Scholarship." (2010).
[44] Alvarado, Rafael C. (2012). The Digital Humanities Situation. Debates in the Digital Humanities, 50-55, University of Minnesota Press, https://doi.org/10.5749/minnesota/9780816677948.003.0005

[45] Bolina, M.: Yewno Discover. Nordic Journal of Information Literacy in Higher Education 11(1) (2019)
[46] Esko Ikkala, Eero Hyvönen, Heikki Rantala, and Mikko Koho. 2021. Sampo-UI: A full stack JavaScript framework for developing semantic portal user interfaces. Semantic Web (2021), 1–16.
[47] Amit Kumar and Marc Spaniol. Annotag: Concise content annotation via lod tags derived from entity-level analytics. In Gerd Berget, Mark Michael Hall, Daniel Brenn, and Sanna Kumpulainen, editors, Linking Theory and Practice of Digital Libraries, pages 175–180. Springer International Publishing, 2021.
[48] Butuc, M. G. (2009). Semantically enriching content using opencalais. Editia, 9, 77-88.
[49] Andrea Loreggia, Simone Mosco, and Alberto Zerbinati. Sentag: A web-based tool for semantic annotation of textual documents. In Gerd Berget, Mark Michael Hall, Daniel Brenn, and Sanna Kumpulainen, editors, ThirtySixth AAAI Conference on Artificial Intelligence. AAAI Press, June 2022.
[50] Simon, Rainer et al. "Linked Data Annotation Without the Pointy Brackets: Introducing Recogito 2." Journal of Map Geography Libraries 13 (2017): 111 - 132.
[51] P.N. Mendes, M. Jakob, A. García-Silva and C. Bizer, DBpedia spotlight: Shedding light on the Web of Documents, in: International Conference on Semantic Systems (I-Semantics), C. Ghidini, A.-C. Ngonga Ngomo, S.N. Lindstaedt and T. Pellegrini, eds, ACM, 2011, pp. 1–8. doi:10.1145/2063518.2063519.
[52] Maria Nisheva Pavlova and Asen Alexandrov. 2020. Extending the GLOBDEF framework with support for semantic enhancement of various data formats. International Journal of Metadata, Semantics and Ontologies 14, 2 (2020), 158–158.
[53] J. Hoffart, M.A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal,M. Spaniol, B. Taneva, S. Thater and G. Weikum, Robust disambiguation of named entities in text, in: Empirical Methods in Natural Language Processing (EMNLP), ACL, 2011,pp. 782–792.
[54] Giannopoulos, G., Bikakis, N., Dalamagas, T., Sellis, T. (2010, May). GoNTogle: a tool for semantic annotation and search. In Extended Semantic Web Conference (pp. 376-380). Springer, Berlin, Heidelberg.
[55] Gangemi, A. (2013, May). A comparison of knowledge extraction tools for the semantic web. In Extended semantic web conference (pp. 351-366). Springer, Berlin, Heidelberg.