

Research Strategy document

Search strategy

Step 1:

We started with meeting with the supervisors about the subject. During this meeting, the supervisors enlightened us about what they expected from us and gave us some papers and directions to start from. We also planned future meetings and decided to meet every week again.

These papers are the following:

1. R&Wbase: Git for triples [\[1\]](#).
2. Comparing a polling and push-based approach for live Open Data interfaces [\[2\]](#).
3. Geospatial partitioning of open transit data [\[3\]](#).

Other subject they told us to read about are the following:

1. [Kafka](#)
2. [Ostrich](#)
3. [Event streams](#)

Step 2

After the first meeting, we started reading in into the subject by reading the provided paper and subjects (from step 1). This provided us with some basic knowledge about the subject. During the following meeting, we discussed our original search strategy. We started by discussing what technologies seemes interesting for the target of the research. We decided that we would look into the following terms:

1. Kafka
2. Git based
3. MQTT
4. DNS
5. Decentralization
6. Azure service bus
7. Linked data streaming
8. Government Open Data

These terms will lead to the actual papers that are found in the following steps.

Step 3

1. Search "mqtt event streaming" on ResearchGate -> atmoko2017.
2. Search on "DNS based data publishing", "DNS based data" on Google Scholar -> resulted in [\[4\]](#) and [\[5\]](#).
3. Search "messaging protocols" on ieeexplore -> Luzuriaga2015.
4. Search "service bus" on Azure-Docs GitHub repository -> <https://github.com/MicrosoftDocs/azure-docs/blob/master/articles/service-bus-messaging/service-bus-messaging-overview.md>
5. Search "DNS update propagation" on ieeexplore -> Parwez2010.

6. Search "DNS-over-HTTPS" on ResearchGate -> Hrushak2020, Borgolte2019HowDO.
7. Search "event streams linked data" on regular Google search -> resulted in [8] and [9].
8. Search "linked data querying" on Google Scholar -> resulted in [10].
9. Search "linked data git" on Google Scholar -> resulted in [11].
10. Search "linked data git" on Google Scholar -> resulted in [12].
11. Search "linked data git" on Google Scholar -> resulted in [15].
12. Search "linked data git" on Google Scholar -> resulted in [16].
13. Search "linked data git" on Google Scholar -> resulted in [17].
14. Search "git versioning of datasets" on Google Scholar -> resulted in [13].
15. Search "linked data stream" on Google Scholar -> resulted in [14].
16. Search "Government Big Open Data" on Web of Science -> resulted in [20].\
17. Search "open data updates" on Web of Science -> resulted in [24].

Step 4

Forward and backward tracking of given and found papers

Backward tracking from Hrushak2020 -> RFC8484.

Use Google Scholar to research papers citing [1] -> resulted in [6] Use Google Scholar to research papers citing [10] -> resulted 10.1007/978-3-319-25007-6_38 Use Google Scholar to research papers citing [18] -> resulted in [19]. Use Google Scholar to research papers cited by found papers (backwards) -> No useful results

Research main contributors of Apache Kafka -> Research of the contributors is not very useful, but found [7].

Use Web of Science to research papers cited in [20] -> resulted in [21]

Use Web of Science to research papers cited in [21] -> resulted in [22]

Use Web of Science to research papers cited in [22] -> resulted in [23]

List of articles of interest

Will certainly appear in the survey paper

- [1] R&wbase proposes a git-based file structure and organisation to allow versioning linked data.
- [3] The paper tries to address the problem of storing too large datasets for (in this case) public transit data. The proposed solution consists of fragmenting the data and making use of linked data, as to limit the amount of data needed to resolve a query.
- [1000] This paper describes the application of MQTT as a communication protocol for IoT devices. We can use this to send updates to clients with very low latency and high performance.
- [10]
- [14] This paper is about C-SPARQL, an extension on top of SPARQL. C-SPARQL makes it possible to query over a stream of RDF triples. Since it is an extension on top of SPARQL, already existing functionalities of SPARQL can be used. The same goes for the SPARQL syntax. The RDF stream is defined as an ordered sequence of pairs, where each pair is made of an RDF triple and its timestamp. Especially the part about how the RDF stream is defined is usefull.
- [24] This paper discribes an approach to updates caches of open data using an application-aware change prioritization approach.

Will most likely appear in the survey paper

- [\[1001\]](#) Exploration of different MoM techniques. Instead of only using MQTT, this paper also presents AMQP which allows unlimited message lengths, as well as storing messages
- [\[6\]](#) The paper presents a new version controlled way of storing and working with Linked Data. The paper also mentions related work such as R&Wbase: git for triples (by P. Colpaert e.a.) and claims this to be an improvement.
- [\[1003\]](#) Instead of running DNS over UDP as is the case today, this paper proposes to use DNS over HTTPS. This could be beneficial, because we can make use of existing caching solutions provided by the HTTP standard.
- [\[9\]](#) Document proposes another way of publishing event streams as linked data, using their own defined entities for events and such. This corresponds to the TREE specification and it could be interesting to compare.
- [\[1004\]](#) A new way to use DNS imposes new privacy and security challenges to tackle. This paper gives an outline of the main concerns.
- [\[11\]](#) The paper is about versioning linked data. this builds on top of the git stack.
- [\[12\]](#) The paper is doing versioning of RDF data using a combination of SPARQL 1.1 and Git.
- [\[13\]](#) The paper is about comparing versions of a git versioned quad store by using Quit diff.
- [\[20\]](#) This paper gives an overview of current data distribution and interpretation mechanisms used by Governments.
- [\[21\]](#) This paper is about the current need and use of intermediaries on government open data.
- [\[23\]](#) This paper gives an overview of the existing state of the art in open data.

Will not likely appear in the survey paper

[\[2\]](#) [\[1002\]](#) [\[7\]](#) [\[8\]](#) [\[15\]](#) [\[16\]](#) [\[17\]](#) [\[1005\]](#) [\[1006\]](#) [\[18\]](#) [\[19\]](#) [\[22\]](#)

Will not appear in the survey paper

[\[4\]](#) [\[5\]](#)

References

[\[1\]](#) Vander Sande, M., Colpaert, P., Verborgh, R., Coppens, S., Mannens, E., & Van de Walle, R. (2013). R&Wbase: git for triples. LDOW, 996.

@article{sande2013r, title={R&wbase: git for triples}, author={Sande, MV and Colpaert, P and Verborgh, R and Coppens, S and Mannens, E and de Walle, RV}, journal={LDOW}, volume={996}, year={2013} }

[\[2\]](#) Van de Vyvere, B., Colpaert, P., & Verborgh, R. (2020, June). Comparing a polling and push-based approach for live Open Data interfaces. In International Conference on Web Engineering (pp. 87-101). Springer, Cham.

@inproceedings{van2020comparing, title={Comparing a polling and push-based approach for live Open Data interfaces}, author={Van de Vyvere, Brecht and Colpaert, Pieter and Verborgh, Ruben}, booktitle={International Conference on Web Engineering}, pages={87--101}, year={2020}, organization={Springer} }

[\[3\]](#) Delva, H., Rojas, J. A., Vandenberghe, P. J., Colpaert, P., & Verborgh, R. (2020, June). Geospatial Partitioning of Open Transit Data. In International Conference on Web Engineering (pp. 305-320). Springer, Cham.

@inproceedings{delva2020geospatial, title={Geospatial Partitioning of Open Transit Data}, author={Delva, Harm and Rojas, Julián Andrés and Vandenberghe, Pieter-Jan and Colpaert, Pieter and Verborgh, Ruben},

booktitle={International Conference on Web Engineering}, pages={305--320}, year={2020}, organization={Springer} }

[1000] @article{atmoko2017, author = {Atmoko, Rachmad and Riantini, R and Hasin, M}, year = {2017}, month = {05}, pages = {012003}, title = {IoT real time data acquisition using MQTT protocol}, volume = {853}, journal = {Journal of Physics: Conference Series}, doi = {10.1088/1742-6596/853/1/012003} }

[4] Spektor, D., & Cleaver, B. A. (2014). U.S. Patent No. 8,689,280. Washington, DC: U.S. Patent and Trademark Office.

[5] Bogner, E. (2011). U.S. Patent No. 7,987,291. Washington, DC: U.S. Patent and Trademark Office.

[1001] @INPROCEEDINGS{Luzuriaga2015, author={J. E. {Luzuriaga} and M. {Perez} and P. {Boronat} and J. C. {Cano} and C. {Calafate} and P. {Manzoni}}, booktitle={2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)}, title={A comparative evaluation of AMQP and MQTT protocols over unstable and mobile networks}, year={2015}, volume={}, number={}, pages={931-936}, doi={10.1109/CCNC.2015.7158101}}

[6] Graube, M., Hensel, S., & Urbas, L. (2014). R43ples: Revisions for triples. In Proceedings of the 1st Workshop on Linked Data Quality co-located with 10th International Conference on Semantic Systems (SEMANTICS 2014).

[1002] @INPROCEEDINGS{Parwez2010, author={M. R. {Parwez} and M. {Akbar} and S. {Haider} and M. S. {Javaid}}, booktitle={2010 2nd IEEE International Conference on Information Management and Engineering}, title={DNS propagation delay: An effective and robust solution using authoritative response from non-authoritative server}, year={2010}, volume={}, number={}, pages={150-153}, doi={10.1109/ICIME.2010.5477485}}

[7] <https://github.com/guozhangwang/awesome-streaming>

[1003] @inproceedings{Hrushak2020, author = {Hrushak, Serhii and Pavlenko, Cynthia}, year = {2020}, month = {04}, pages = {}, title = {Advantages of DNS-over-HTTPS over DNS}, doi = {10.30837/IVcsitic2020201356} }

[8] Tommasini, R., Ragab, M., Falcetta, A., Della Valle, E., & Sakr, S. (2019). Bootstrapping the Publication of Linked Data Streams. In ISWC Satellites (pp. 29-32).

[9] Harth, A., Stühmer, R. Publishing Event Streams as Linked Data. <https://km.aifb.kit.edu/sites/lodstream/>

[1004] @inproceedings{Borgolte2019HowDO, title={How DNS over HTTPS is Reshaping Privacy, Performance, and Policy in the Internet Ecosystem}, author={Kevin Borgolte and T. Chattopadhyay and N. Feamster and Mihir Kshirsagar and Jordan Holland and Austin Hounsel and Paul Schmitt}, year={2019} }

[10] Verborgh, R., Vander Sande, M., Colpaert, P., Coppens, S., Mannens, E., & Van de Walle, R. (2014, April). Web-Scale Querying through Linked Data Fragments. In LDOW.

[11] Arndt, N., Naumann, P., & Marx, E. (2017, May). Exploring the Evolution and Provenance of Git Versioned RDF Data. In MEPDaW/LDQ@ ESWC (pp. 12-27).

@inproceedings{arndt2017exploring, title={Exploring the Evolution and Provenance of Git Versioned RDF Data.}, author={Arndt, Natanael and Naumann, Patrick and Marx, Edgard}, booktitle={MEPDaW/LDQ@ ESWC}, pages={12--27}, year={2017} }

[12] Arndt, N., Radtke, N., & Martin, M. (2016, September). Distributed collaboration on rdf datasets using git: Towards the quit store. In Proceedings of the 12th International Conference on Semantic Systems (pp. 25-32).

@inproceedings{arndt2016distributed, title={Distributed collaboration on rdf datasets using git: Towards the quit store}, author={Arndt, Natanael and Radtke, Norman and Martin, Michael}, booktitle={Proceedings of the 12th International Conference on Semantic Systems}, pages={25--32}, year={2016} }

[13] Arndt, N., & Radtke, N. (2016, September). Quit diff: Calculating the delta between rdf datasets under version control. In Proceedings of the 12th International Conference on Semantic Systems (pp. 185-188).

@inproceedings{arndt2016quit, title={Quit diff: Calculating the delta between rdf datasets under version control}, author={Arndt, Natanael and Radtke, Norman}, booktitle={Proceedings of the 12th International Conference on Semantic Systems}, pages={185--188}, year={2016} }

[14] Barbieri, D. F., & Valle, E. D. (2010, April). A proposal for publishing data streams as linked data. In Linked Data on the Web Workshop.

@inproceedings{barbieri2010proposal, title={A proposal for publishing data streams as linked data}, author={Barbieri, Davide F and Valle, ED}, booktitle={Linked Data on the Web Workshop}, year={2010} }

[15] Meroño-Peñuela, A., & Hoekstra, R. (2017, May). SPARQL2Git: transparent SPARQL and linked data API curation via Git. In European Semantic Web Conference (pp. 143-148). Springer, Cham.

@inproceedings{merono2017sparql2git, title={SPARQL2Git: transparent SPARQL and linked data API curation via Git}, author={Meroño-Peñuela, Albert and Hoekstra, Rinke}, booktitle={European Semantic Web Conference}, pages={143--148}, year={2017}, organization={Springer} }

[16] Kubitza, D. O., Böckmann, M., & Graux, D. (2019, October). SemanGit: A linked dataset from git. In International Semantic Web Conference (pp. 215-228). Springer, Cham.

@inproceedings{kubitza2019semangit, title={SemanGit: A linked dataset from git}, author={Kubitza, Dennis Oliver and Böckmann, Matthias and Graux, Damien}, booktitle={International Semantic Web Conference}, pages={215--228}, year={2019}, organization={Springer} }

[17] Meroño-Peñuela, A., & Hoekstra, R. (2016, May). grlc makes GitHub taste like linked data APIs. In European Semantic Web Conference (pp. 342-353). Springer, Cham.

@inproceedings{merono2016grlc, title={grlc makes GitHub taste like linked data APIs}, author={Meroño-Peñuela, Albert and Hoekstra, Rinke}, booktitle={European Semantic Web Conference}, pages={342--353}, year={2016}, organization={Springer} }

[1005] @InProceedings{10.1007/978-3-319-25007-6_38, author="Kuhn, Tobias and Chichester, Christine and Krauthammer, Michael and Dumontier, Michel", editor="Arenas, Marcelo and Corcho, Oscar and Simperl, Elena and Strohmaier, Markus and d'Aquin, Mathieu and Srinivas, Kavitha and Groth, Paul and Dumontier, Michel and Heflin, Jeff and Thirunarayan, Krishnaprasad and Thirunarayan, Krishnaprasad and Staab, Steffen", title="Publishing Without Publishers: A Decentralized Approach to Dissemination, Retrieval, and Archiving of Data", booktitle="The Semantic Web - ISWC 2015", year="2015", publisher="Springer International Publishing", address="Cham", pages="656--672", isbn="978-3-319-25007-6" }

[1006] @techreport{RFC8484, author = {P. Hoffman and P. McManus}, title = {DNS Queries over HTTPS (DoH)}, howpublished = {Internet Requests for Comments}, type = {RFC}, number = {8484}, year = {2018},

month = {October}, issn = {2070-1721}, publisher = {RFC Editor}, institution = {RFC Editor}, }

[18] Verborgh, R., Hartig, O., De Meester, B., Haesendonck, G., De Vocht, L., Vander Sande, M., ... & Van de Walle, R. (2014, October). Querying datasets on the web with high availability. In International Semantic Web Conference (pp. 180-196). Springer, Cham.

[19] Folz, P., Skaf-Molli, H., & Molli, P. (2016, May). CyCLaDEs: a decentralized cache for Linked Data Fragments. In ESWC: Extended Semantic Web Conference.

[20] Lnenicka, M., Komarkova, J. (2019, June). Developing a government enterprise architecture framework to support the requirements of big and open linked data with the use of cloud computing

[21] van Schalkwyk, F., Willmers, M., McNaughton M. (2015, October). Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem

[22] Dwivedi, Y., Janssen, M., Slade, E., Rana, N., Weerakkody, V., Millard, J., Hidders, J., Snijders, D., (2016, July). Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural modelling

[23] Hossain, M., Dwivedi, Y., Rana, N. (2016, April). State-of-the-art in open data research: Insights from existing literature and a research agenda

[24] U. Akhtar et al., "Change-Aware Scheduling for Effectively Updating Linked Open Data Caches," in IEEE Access, vol. 6, pp. 65862-65873, 2018, doi: 10.1109/ACCESS.2018.2871511.