RADON results for OAEI 2017

Kevin Dreßler[2], Mohamed Ahmed Sherif[1], and Axel-Cyrille Ngonga Ngomo[1]

[1] Paderborn University, Data Science Group, Pohlweg 51, D-33098 Paderborn, Germany
E-mail: {firstname.lastname}@upb.de
[2] Department of Computer Science, University of Leipzig, 04109 Leipzig, Germany
E-mail: {lastname}@informatik.uni-leipzig.de

Abstract. Datasets containing billions of geospatial resources are increasingly being represented according to the Linked Data principles. RADON is an efficient solution for the discovery of topological relations between such geospatial resources according to the DE9-IM standard. RADON uses a sparse space tiling index in combination with minimum bounding boxes to reduce the computation time of topological relations. In this paper, we present the participation of RADON in the OAEI 2017 campaign. The OAEI results show that RADON outperforms the other state of the art significantly in most of the cases.

1 Presentation of the system

RADON is a time-efficient link discovery algorithm for topological relations between geospatial resources, implemented within LIMES [3].

Given two sets of RDF resources S and T and a relation R, the goal of link discovery is to find the *mapping* $M = \{(s,t) \in S \times T : R(s,t)\}$. RADON enables the time-efficient discovery of all topological relations that can be defined in terms of the DE-9IM standard [1]. In order to achieve time-efficiency, two optimization techniques are utilized: *optimized sparse space tiling* on the dataset level and *Minimum Bounding Box (MBB)-based filtering* on the resource level.

In the following, we introduce the basic concepts needed to understand RADON before we outline the aforementioned optimization techniques. More detailed explanations can be found in [5].

The Minimum Bounding Box (MBB) of a geometry g in n dimensions is the rectangular box with the smallest measure (area, volume, or hypervolume in higher dimensions) within which all points of g lie. Another term for MBB is *envelope*.

Space tiling is a technique for indexing spatial data, where n-dimensional affine spaces are split into any number of hyperrectangles with edge lengths ℓ_i and granularity factors $\Delta_i = (\ell_i)^{-1}$ where $i \in \{1, \ldots, n\}$. These hyperrectangles can then be addressed using vectors from $\mathbb{N}^n$, which allows for various optimizations.

1.1 Optimized Sparse Space Tiling

The goal of the *optimized sparse space tiling* is to generate an index I for mapping all geometries $s \in S, t \in T$ to sets of hyperrectangles. For the sake of clarity, the following description focuses on the two-dimensional case. As a first step, we use a heuristic to get good granularity factors for both latitude and longitude dimensions $(\Delta_\varphi, \Delta_\lambda)$. Then, we apply space tiling, in which we map a geometry g to the set of hyperrectangles over which its MBB spans. To implement this idea, we insert a reference to g into all those hyperrectangles, that are realized as entries of a `HashMap`. To optimize (i.e. sparsify) the generated index, we start by computing estimated total hypervolumes (ETH) of the datasets S and T. We first index the dataset with the smaller ETH for each resource of the other dataset. We then add only to I the subset of resources from the second dataset which shares the same hyperrectangles from the first dataset resources contained in I. Using this technique together with the `HashMap` implementation of the hyperrectangle index significantly reduces the size of the generated data structure and consequently also the time to traverse it.

1.2 MBB-based Filtering

After the *optimized sparse space tiling* step described above, we traverse the generated index, visiting one hyperrectangle at a time. As a consequence of our approach, each generated hyperrectangle contains references to at least one geometry from each dataset. For each pair (s, t) of geometries, where $s \in S$ and $t \in T$, we then employ a filtering step before actually triggering the potentially expensive (in cases of large geometries) computation that checks if the given relation holds. Let $\square(g)$ denote the MBB of geometry g. The filtering step leverages the fact that $\neg r(\square(s), \square(t)) \Rightarrow \neg r(s, t)$ holds for every relation r, where one geometry has no interior or boundary points in the exterior of the other geometry, i.e. $s \subseteq t$ or $t \subseteq s$. For these relations, we can return false and skip further computations, iff the geometries MBB's do not satisfy the relation.

2 Adaptations made for the evaluation

No specific adaptations were made to the original RADON algorithm [5], we only provide a Java `SystemAdapter` according to the campaign guidelines[3]. The final RADON Java `SystemAdapter` source code is available online in the project website[4].

3 Evaluation Results

RADON has been evaluated only in the *Hobbit Link Discovery Track Task 2 (Spatial)*. The basic idea behind this task was to measure how well the systems can identify DE-9IM (Dimensionally Extended nine-Intersection Model) topological relations. The supported spatial relations were: *Disjoint, Touches, Contains/Within, Covers/CoveredBy,*

[3] `https://goo.gl/cWmZ5P`
[4] `https://goo.gl/awkvvo`

Intersects, Crosses, Overlaps. The geospatial resources traces were represented in Well-known text (WKT) format as `LineStrings`.

Given two sets of `LineString` geometries S and T and a DE-9IM topological relation R, the participants were assigned the task of retrieving the mapping $M = \{(s,t) \in S \times T : R(s,t)\}$. All the systems were tested against two datasets: (1) the *sandbox* dataset, with a scale of 10 instances, and (2) the *mainbox* dataset with a scale of 2K instances.

The other participants to this task in addition to RADON were AGREEMENTMAKERLIGHT (AML), ONTOIDEA, and SILK. The systems were judged on the basis of precision, recall, F-Measure and run time. The final results are shown in Table 1 and Figures 1 and 2. Note that we are only presenting the time performance and not precision, recall and F-Measure, as all were equal to 1.0 except ONTOIDEA TOUCHES and OVERLAPS which is equal to 0.99.

From these results we can see that, while RADON performs in the middle field of the the sandbox dataset, RADON outperforms the other participants on most relations for the sandbox dataset. Notably, the optimization described in Section 1.2 speeds up the relations *Equals, Contains, Within, Covers* and *CoveredBy* significantly in comparison to the remaining relations. The differences in performance between *Touches, Intersects*, where AML outperforms RADON, and *Overlaps* cannot be explained from an implementation point of view, as these three relations share the exact optimizations. However, due to the datasets consisting exclusively of `LineStrings`, it is apparent that *Touches* and *Intersects* are much more likely to hold between any two geometries than *Overlaps*. Therefore, the benchmarks on these relations are the hardest in this task.

4 Conclusion

We priefly presented RADON, an approach for rapid discovery of topological relations among geo-spatial resources. To achieve a high scalability, RADON combines space tiling, minimum bounding box approximation and a sparse index. The presented evaluation during the OAEI 2017 showed that, in addition to being complete and correct (i.e. achieving an F-Measure of 1.0), RADON also outperforms the other participating systems in most of the cases. In future work, we aim to apply the particle-swarm-optimization load balancing approaches [6]. To improve the performance of RADON on high resolution datasets, i.e. datasets whose containing geometries consist of a large set of points, we will optimize the computation of relation checks. In order to further reduce the amount of computations, we will consider adaptive granularity factors, i.e. granularity factors as functions of latitude and longitude. In addition, we aim to combine RADON with the machine learning approaches already implemented in LIMES such as the WOMBAT [4] algorithm. Finally, we will consider the discovery of *temporospatial* relations, by integrating the AEGLE[2] algorithm with the RADON approach.

Acknowledgments

This work has been supported by the eurostars project SAGE (GA no. E!10882), the H2020 projects SLIPO (GA no. 731581) and HOBBIT (GA no. 688227) as well as the

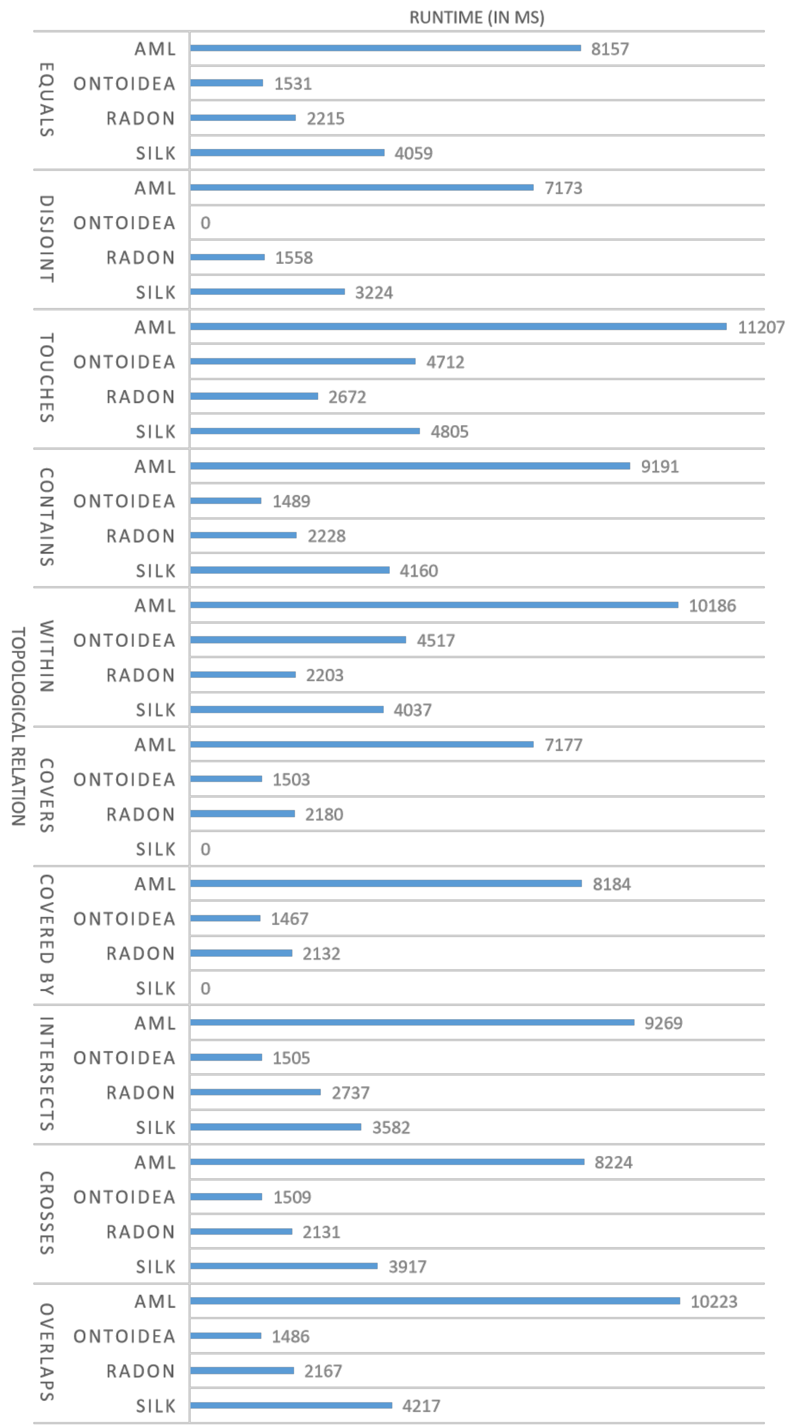

Fig. 1. Runtime comparison for *Sandbox* dataset

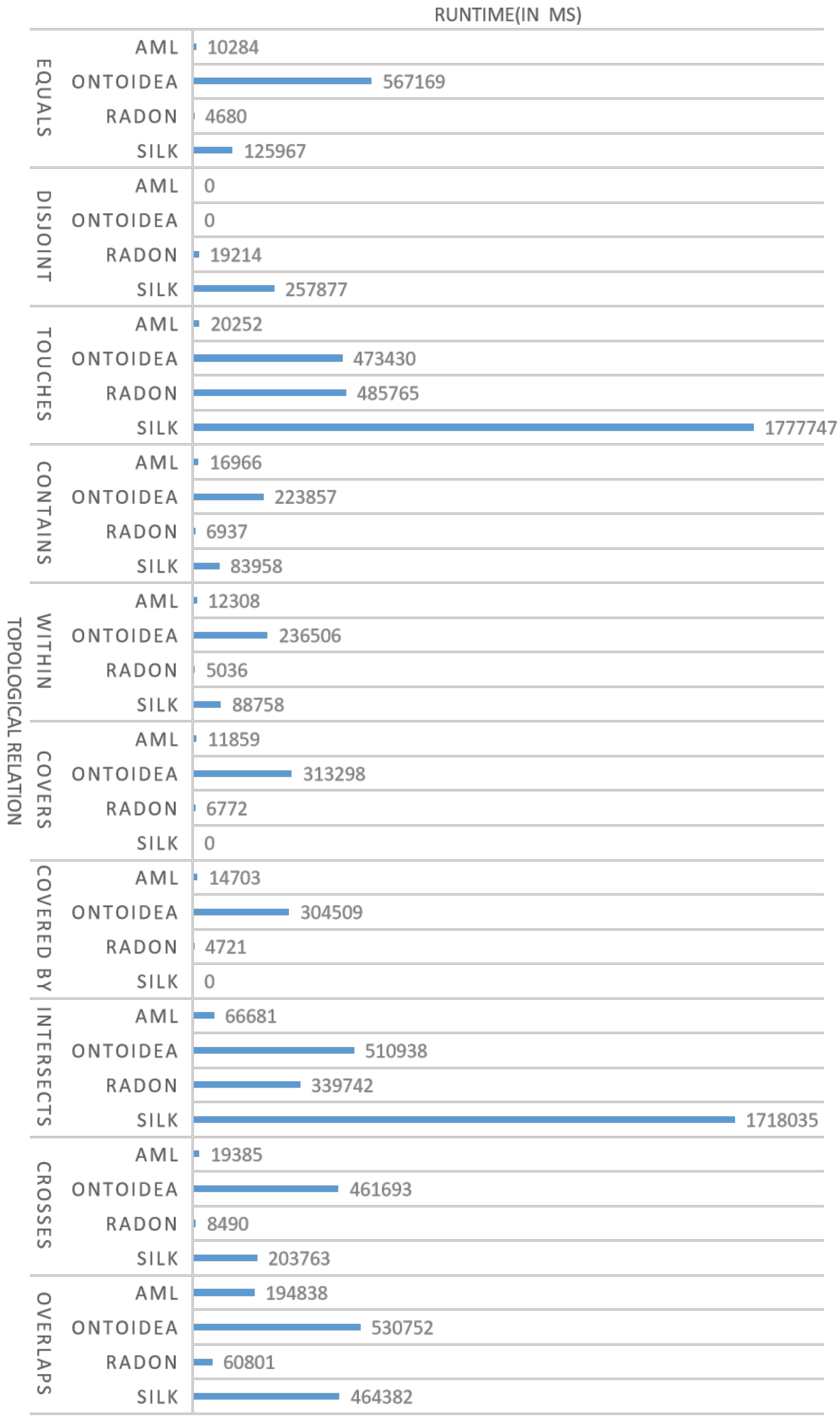

Fig. 2. Runtime comparison for *Mainbox* dataset

Table 1. Hobbit link discovery task evaluation results for all participants. Note that we used — for systems which were not participating in the specified sub-task and × for systems that exceeded the time limit.

Relation	System	Sandbox	Mainbox
Equals	AML	8157	10284
	OntoIdea	1531	567169
	Radon	2215	4680
	Silk	4059	125967
Disjoint	AML	7173	×
	OntoIdea	—	—
	Radon	1558	19214
	Silk	3224	257877
Touches	AML	11207	20252
	OntoIdea	4712	473430
	Radon	2672	485765
	Silk	4805	1777747
Contains	AML	9191	16966
	OntoIdea	1489	223857
	Radon	2228	6937
	Silk	4160	83958
Within	AML	10186	12308
	OntoIdea	4517	236506
	Radon	2203	5036
	Silk	4037	88758
Covers	AML	7177	11859
	OntoIdea	1503	313298
	Radon	2180	6772
	Silk	—	—
CoveredBy	AML	8184	14703
	OntoIdea	1467	304509
	Radon	2132	4721
	Silk	—	—
Intersects	AML	9269	66681
	OntoIdea	1505	510938
	Radon	2737	339742
	Silk	3582	1718035
Crosses	AML	8224	19385
	OntoIdea	1509	461693
	Radon	2131	8490
	Silk	3917	203763
Overlaps	AML	10223	194838
	OntoIdea	1486	530752
	Radon	2167	60801
	Silk	4217	464382

DFG project LinkingLOD (project no. NG 105/3-2) and the BMWI Project GEISER (project no. 01MD16014E).

References

1. E. Clementini, J. Sharma, and M. J. Egenhofer. Modelling topological spatial relations: Strategies for query processing. *Computers & graphics*, 18(6):815–822, 1994.
2. K. Georgala, M. A. Sherif, and A.-C. N. Ngomo. An efficient approach for the generation of allen relations. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI) 2016, The Hague, 29. August - 02. September 2016*, 2016.
3. A.-C. Ngonga Ngomo and S. Auer. Limes - a time-efficient approach for large-scale link discovery on the web of data. In *Proceedings of IJCAI*, 2011.
4. M. Sherif, A.-C. Ngonga Ngomo, and J. Lehmann. WOMBAT - A Generalization Approach for Automatic Link Discovery. In *14th Extended Semantic Web Conference, Portorož, Slovenia, 28th May - 1st June 2017*. Springer, 2017.
5. M. A. Sherif, K. Dreßler, P. Smeros, and A.-C. N. Ngomo. Radon-rapid discovery of topological relations. In *AAAI*, pages 175–181, 2017.
6. M. A. Sherif and A.-C. N. Ngomo. An optimization approach for load balancing in parallel link discovery. In *Proceedings of the 11th International Conference on Semantic Systems*, SEMANTICS '15, pages 161–168, New York, NY, USA, 2015. ACM.