Towards Duplicate Detection for Situation Awareness Based on Spatio-Temporal Relations*

Norbert Baumgartner[1], Wolfgang Gottesheim[2], Stefan Mitsch[2], Werner Retschitzegger[2], and Wieland Schwinger[2]

[1] team Communication Technology Mgt. Ltd., Goethegasse 3, 1010 Vienna, Austria
[2] Johannes Kepler University Linz, Altenbergerstr. 69, 4040 Linz, Austria

Abstract. Systems supporting situation awareness typically integrate information about a large number of real-world objects anchored in time and space provided by multiple sources. These sources are often characterized by identical, incomplete, and even contradictory information. Because of that, duplicate detection methods are of paramount importance, allowing to explore whether or not information concerns one and the same real-world object. Although many such duplicate detection methods exist, a recent survey revealed that the characteristics of situation awareness—highly dynamic and vague information, which is often available in qualitative form only—are not supported sufficiently well. This paper proposes concepts for qualitative duplicate detection to cope with these key issues of situation awareness based on spatio-temporal relations between objects.

1 Introduction

Situation awareness [17] is gaining more and more importance as a way to help human operators cope with information overload in large-scale control systems [6], like, e. g., encountered in the domain of road traffic management [26]. Systems supporting situation awareness typically deal with a *vast amount of information* about a large number of *real-world objects* anchored in *time and space* provided by *multiple sources*. These sources are often characterized by heterogeneous formats and, most crucial, contain identical, incomplete, and often even contradictory information [30]. Besides having to resolve structural heterogeneities at schema level, the data itself has to be fused into a single consistent form at instance level [8].

Characteristics of duplicates in situation awareness. As a major prerequisite for the latter task, duplicate detection methods are of paramount importance, allowing to explore whether or not information having, e. g., different origins or different observation times, concern one and the same real-world object. As was pointed out in a recent survey [5], existing duplicate detection methods, however, fail to consider the characteristics of situation awareness with respect to duplicate detection. In particular, the underlying data about real-world objects (e. g., a traffic jam) is often highly dynamic and vague. This entails that reliable numerical values are hard to obtain, thus making *qualitative*

* This work has been funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) under grant FIT-IT 819577.

duplicate detection approaches better suited than quantitative ones [14]. Current qualitative duplicate detection methods, however, are highly domain-specific, like e. g., such being proposed for detecting similar trajectories of moving objects in a road network [23], and measure similarity only with respect to a single spatial aspect, like distances between points of interest, at most taking time as additional dimension into account.

Duplicate detection on the basis of qualitative spatio-temporal relations. To overcome these limitations, we base upon the observation of Keane et al. [25], who state that humans generally tend to base their similarity judgement on (multiple) relation similarities, if present. Such relations can be expressed by employing relation calculi, each of them focusing on a certain spatio-temporal aspect, like mereotopology [31], orientation [15], or direction [29]. These calculi are often formalized by means of Conceptual Neighborhood Graphs (CNGs, [19]), imposing constraints on the transitions between relations. By that, CNGs are an important notion for modeling continuously varying processes [28], and are adopted in, e. g., qualitative simulation [12], prediction [7], tracking moving objects [34], or agent control [16]. We propose a duplicate detection approach on the basis of similarity measures derived from relation calculi and their CNGs, comprising *rule-based* similarity measures defined by domain experts, which are accompanied by *distance-based* ones. With rule-based similarity measures, domain experts can express their knowledge for detection of identical information (e. g., two objects are duplicates, if they are in the same-type- and equal-region-relationship). Complementary, with distance-based similarity measures duplicates in incomplete and contradictory information can be detected as well (e. g., two traffic jams being in `VeryFar` relationship from distance calculus are less likely duplicates then such being `VeryClose`). The applicability of this approach is demonstrated on the basis of real-world traffic data collected by the Austrian Highways Agency ASFINAG.

Structure of the paper. In the next section, we summarize the characteristics of situation awareness with respect to duplicate detection by means of examples from the domain of road traffic management. In Section 3, we propose concepts for duplicate detection being applicable in presence of such characteristics. We then evaluate these concepts in Section 4. Finally, we provide an overview of related work in Section 5, before we conclude the paper in Section 6 by indicating future work.

2 Motivating Example

Road traffic management systems, being responsible for, e. g., improving traffic flow and ensuring safe driving conditions, are a typical application domain for situation awareness, aiming at assisting human operators by providing reliable qualitative information, e. g., "a wrong-way driver heading towards a traffic jam". In our previous work [6] we introduced a framework[3] for building situation-aware systems on the basis of a domain-independent ontology representing qualitative information about the system under control. In the framework's ontology, *objects* and *relations* between them are aggregated into *situations*, whereas objects can either be real-world objects with a physical identity (e. g., a tunnel) or reified representations of events (e. g., a traffic

[3] This framework was used to build a prototype supporting human operators, thereby showing the feasibility of our approach.

jam). Information about such objects and relations stems from various sources being maintained independently, which, naturally, leads to duplicates in the form of identical, incomplete, and even contradictory information about the same real-world object. Changes of objects over time (i. e., object evolution), like, e. g., movement on a road, also lead to changes in relations between objects and make it particularly challenging to detect duplicates. In the following, we will exemplify these specific issues.

Let us suppose that a traffic jam builds up on a highway during rush hour, which may lead to a sequence of entries (cf. Fig. 1, 1–5) in the road traffic management system, originating from a call center and from a traffic flow detector. From a chronological

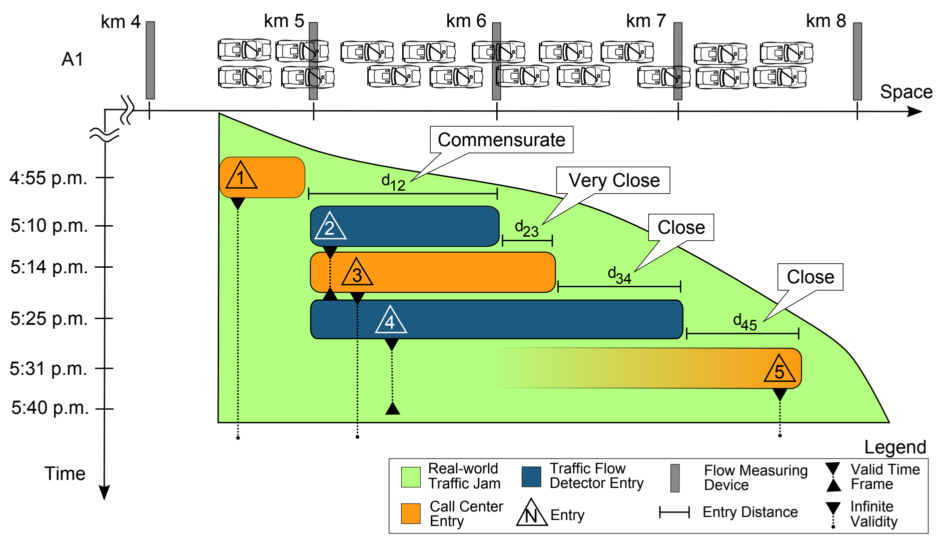

Fig. 1. Information about a traffic jam during rush hour.

point of view, first of all a motorist informs the call center (entry 1). As the traffic jam's starting point is located between two flow measuring devices, it takes a while until the traffic jam has grown to an extent also observed by the traffic flow detector (entry 2). Both the traffic flow detector and the call center subsequently report updated information, as described by entries 3–5. However, motorists located at the end of the traffic jam are less and less able to observe the traffic jam's whole extent, resulting in inexact information about its starting point (entry 3), or even just in information about its end (entry 5).

In order to identify that all of these entries in fact describe one and the same real-world object, existing duplicate detection methods typically compute a *similarity measure* from *selected properties* [8]. Fig. 1 shows an exemplary similarity measure based on the distance between the end positions of traffic jam entries (e. g., d_{23}). Such duplicate detection methods, however, fail if reliable numerical values cannot be provided and if objects dynamically evolve [5], which are both important characteristics of application domains for situation awareness. For example, a "traditional" similarity measure would calculate d_{23} as the quantitative distance between the positions of entry 2 and entry 3, which is not possible if the only information available is, that these positions are *close* to each other. In the following section, we propose concepts to overcome these limitations. Please note that, since we build on our previous work [6], we assume that structural heterogeneitis between data sources have already been resolved.

3 Measuring Similarity of Qualitative Spatio-Temporal Data

A major task in duplicate detection for situation awareness, as exemplified in the previous section, is to measure the similarity of entries being anchored in time and space. In this section, we describe our approach to defining similarity measures on the basis of *spatio-temporal relations*, which are an integral part of describing objects in situations. In contrast to traditional duplicate detection methods based solely on an object's properties, such similarity measures are in accordance with Situation theory of Barwise and Perry [3][4] and exploit the observation of Keane et al. [25], who state that humans generally tend to base their similarity judgement on relation similarities, if present.

In the field of qualitative spatio-temporal reasoning, a number of well-established relation calculi, each of them focusing on a certain aspect, like e. g., mereotopology in the Region Calculus (RCC, [10]) or time in Allen's Time Intervals Algebra [1], structure the possible relationships between objects (e. g., in RCC depicted in Fig. 2, a region occupied by a traffic jam can be `Disrelated` from, `PartiallyOverlapping`, `ProperPart`, or `Equal` to a region of another traffic jam). These calculi are often formalized by means of *conceptual neighborhood graphs* (CNG, cf. [18]), resembling simple state transition diagrams. A CNG represents relations as nodes and the transitions between them as edges, and is arranged consistent with the human perception of similarity, i. e., similar relations are close to each other in a CNG, dissimilar ones are farther apart from each other in terms of the number of transitions in-between (termed *relation distance*). Fig. 2 shows the CNG of RCC, whereby the meaning of each relation is depicted inside the relation's node with circular objects. It can be seen, that the calculus contains one relation, which holds in case two objects are equal (i. e., the identity relation of RCC), three relations for describing somewhat similar objects (`ProperPart`, `PartiallyOverlapping`, and `ProperPartInverse`, because their regions in some form overlap or are contained in each other), and one relation for describing dissimilar objects (`Disrelated`, being farthest from `Equal`). Based on this, in the following we describe how to determine the similarity of entries, on the one hand, using exact definitions given by domain experts in a *rule-based* manner, and, on the other hand, by measuring *relation distances*.

3.1 Rule-Based Similarity Measure

Recently, ontology-driven situation awareness techniques [11], [2] have emerged, providing the vocabulary for describing situations with rules [6], [27]. Such approaches

[4] This notion makes the proposed approach applicable to a wide range of efforts in situation awareness, like e. g., to Kokar's approach [27].

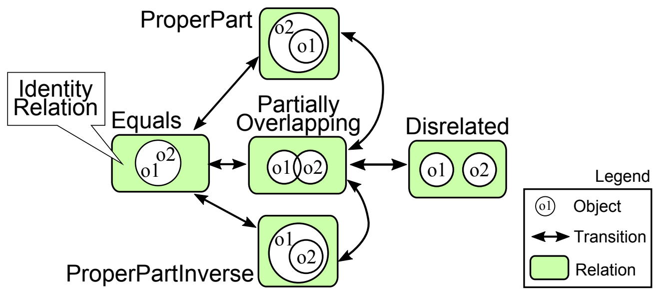

Fig. 2. Conceptual Neighborhood Graph of RCC.

define the necessary pre-requisites for a rule-based duplicate detection approach, allowing domain experts to describe their knowledge about duplicates in the form of exact definitions. In the following example, the rule

$$\begin{aligned} Accident(?a) \wedge Accident(?b) \\ \wedge rcc : Equals(?a, ?b) \wedge allen : Equals(?a, ?b) \end{aligned} \tag{1}$$

defines that, for determining whether two accident entries a and b are duplicates, equality of their regions ($rcc : Equals(?a, ?b)$) as well as their temporal intervals ($allen : Equals(?a, ?b)$, i. e., valid time frames in Fig. 1) is most relevant. Such rules are suitable for detecting identical duplicates, and therefore need to be accompanied with concepts allowing the detection of duplicates in incomplete and contradictory information, as described in the following section.

3.2 Distance-Based Similarity Measure

We can use the distance between relations in a CNG to describe similarity of objects: for this, we identify an *identity relation* in each calculus that is valid if two objects are equal. This identity relation then serves as a reference point for calculating the distance to the relation being actually valid between the two objects. For example, the distance between `Disrelated` and `Equals`—being the identity relation of RCC—is 2 according to the CNG in Fig. 2, whereas `PartiallyOverlapping` and `Equals` are direct neighbors (distance 1). Therefore, if `PartiallyOverlapping` is actually valid between two objects, they are considered to be more similar than two objects being `Disrelated`[5].

Such a very basic one-dimensional computation, however, does not reliably describe the similarity of two objects being characterized by many different real-world aspects. Therefore, we combine calculi describing such different real-world aspects, like mereotopology, orientation, distance, and size from the spatial domain, as well as distance and size from the temporal domain. For combining these calculi, we base upon

[5] Note, that for using a CNG to determine a similarity measure, a calculus needs to be *joint exhaustively and pairwise disjoint (JEPD)*, which means that between any two objects *exactly one* of the calculus' relations is valid.

the work of Schwering [32], which uses a variant of Gärdenfors' *conceptual spaces* [22] to combine different metrics to a similarity measure for relations. In this approach, the similarity of natural-language relation constructs is measured by interpreting the conceptual space as a geometric space delimited by quality dimensions, making it possible to calculate distances within the space. Going beyond the work of Schwering, however, our approach uses such a conceptual space to represent *each relation calculus* on its own quality dimension, and thereby calculates a similarity measure for *entries representing objects*, as depicted in Fig. 3. In this figure, we show the valid relations between

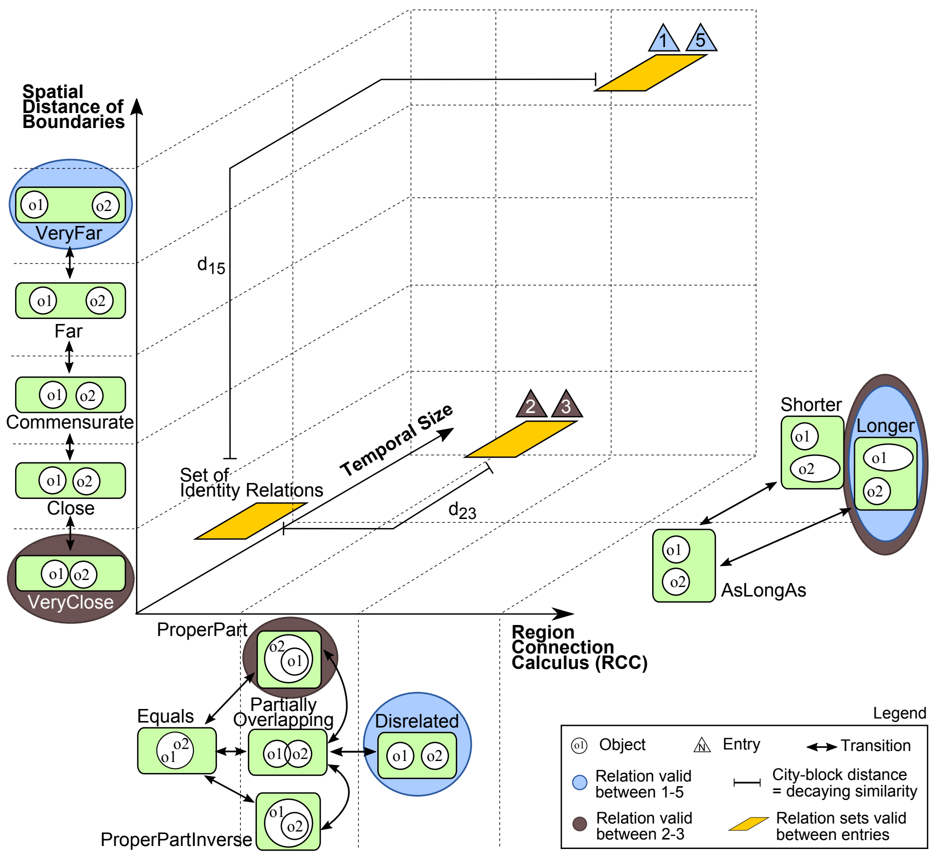

Fig. 3. Entry distances in a conceptual space.

some of the entries from our motivating example. For example, the entries 1 and 5 are `Disrelated` in the quality dimension RCC, their spatial distance is `VeryFar`, and entry 1 is `Longer` in temporal size than entry 5. The relations valid between entry 2 and entry 3 are `ProperPart`, `VeryClose` and `Longer`. The fact that the relation distance defines a metric for each of the quality dimensions allows us to calculate distances between *relation sets* (like, e. g., the relation sets {`Disrelated`, `VeryFar`, Longer}, {ProperPart, VeryClose, Longer}, and {Equals, VeryClose, AsLongAs} depicted in Fig. 3). Particularly interesting is the distance between the set of identity relations and the entries' valid relations, as we define the similarity measure of entries as a decaying function of this distance. For example, the distance d_{15} between the set of identity relations and the relations of entries 1 and 5 indicates that these entries are unlikely to be duplicates. In comparison, the distance d_{23} is smaller, indicating that the entries 2 and 3 are more likely to be duplicates. For calculating such a distance, we can use well-known metrics like Euclidian distance or city-block distance (also known as Manhattan distance) [13]. We follow the suggestion of Johannesson [24] to use the city-block distance, allowing to combine semantically separable dimensions (e. g., time and space), which are predominant in situation awareness.

Since the calculation of relation distances requires comparing every object with every other object, the proposed approach's runtime behaviour can be improved using clustering strategies that reduce the number of object comparison operations. One such strategy could cluster objects by their type, based, e. g., on type subsumption information defined in ontologies, so that only objects having the same type are compared. Other possible strategies include clustering by spatial information, since, e. g., accidents that have occurred on different roads can be assumed to be distinct and, therefore, do not require the calculation of relation distances. Additional improvements are possible by incorporating a domain expert's knowledge formalized in rules as shown in Sect. 3.1 to select only appropriate calculi and identity relations for inclusion in the conceptual space (alternatively, all available calculi could be included), thereby minimizing the number of required comparison operations. Further improvements in terms of accuracy might be achieved by integrating additional qualitative dimensions into the conceptual space and taking knowledge on correlations and subsumptions between calculi into account.

4 Evaluation

For evaluating the proposed concepts, we collaborate with our project partners[6] providing real-world data. These data are reported by multiple sources, like, e. g., traffic flow sensors, road maintenance schedules, and motorists reporting incidents to a call center. The recorded data set used for this evaluation consists of 3,563 distinct road traffic objects, comprising, among others, 778 traffic jams, 819 road works, 1,339 other obstructions, 460 accidents, and 64 weather warnings. In this early stage, we focussed on detecting duplicate traffic jams, and relied on the help of domain experts, who manually examined the recorded data set, resulting in a test data set comprising 35 duplicates among a total of 94 traffic jam messages. It has to be noted, that the presented preliminary evaluation results, although they indicate the applicability of the proposed approach, further need critical observation, as we continuously extend our test data set.

For evaluating duplicate detection we first employed duplicate detection rules specifying that two traffic jam entries are duplicates if the following condition holds: RCC:−

[6] We are currently realizing ontology-driven situation awareness techniques for the domain of road traffic management together with our project partners Heusch/Boesefeldt, a German supplier of road traffic management systems, and the Austrian highways agency ASFINAG.

PartiallyOverlapping $\wedge$ TemporalDistance:Close, cf. Fig. 4 (a). In contrast to that, our approach exploiting relation distances was configured with a two-dimensional conceptual space measuring distances in RCC and Allen's Time Intervals Algebra (allowing duplicates to have a distance of one to each dimension's identity relation), cf. Fig. 4 (b).

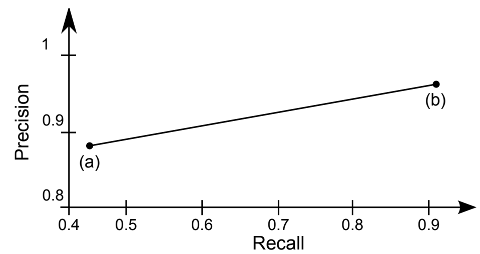

Fig. 4. Precision and recall of proposed concepts.

As the evaluation reveals, duplicate detection relying on rules is only able to detect information exactly satisfying the rule, which leads, on the one hand, to high precision, but on the other hand to many missed duplicates (20, resulting in low recall). In contrast, working with similarity measures based on relation distances holds the risk of interpreting non-duplicates as duplicates (i. e., an increase in *false positives* from two to three), but enables us to detect also duplicates stemming from incomplete or contradictory information (an increase from 15 to 33 true positives, a decrease from 20 to two false negatives, resulting in higher recall). When examining the false positives in detail, we found out that they mainly stem from incomplete temporal information, resulting in entries having no ending time and thus entailing overlapping time intervals. An experienced human operator could substitute such missing temporal information with typical object evolutions, thus avoiding this issue. This observation underpins the hypothesis that object evolution characteristics—as we proposed as an optimization technique for predicting object evolutions [4]—are promising for improving duplicate detection.

5 Related Work

Relevant related work for qualitative duplicate detection in situation awareness can be found in geospatial database research, as well as in research on moving object trajectories.

The survey of Schwering [33] focuses on semantic similarity of geospatial data, thereby emphasizing the importance of appropriate similarity measures for different spatial representation models. One such similarity measure for spatial relations is proposed in [32]. However, this similarity measure is restricted to topologic and distance aspects of relations between lines and regions, and leaves the incorporation of directional relations, or non-spatial relations as further work. We generalize this work to measure similarity between arbitrary *objects* by comparing the relations between them with appropriate *identity relations*.

Similarity analysis of trajectories is an area concerned with comparing the traces of moving objects in time and space [21]. Several such methods exist, like [9], [20], and [23]. However, most of them in fact measure similarity in Euclidian space only (whereas in our domain graphs are more suitable to describe spatial information), and lack support for qualitative spatio-temporal information. The work of Hwang et al. [23] is particularly relevant for our work due to its applicability to road networks. Hwang et al. propose spatio-temporal similarity measures to detect duplicates in trajectories of moving objects. Being based on qualitative information in terms of "points of interest" and "times of interest", the method, however, is not able to consider multiple different spatial aspects, resulting in rudimentary duplicate detection only.

6 Future Work

In our future work we plan to concentrate on two major areas: on the one hand, we aim to improve the detection of inexact duplicates, and on the other hand, we will incorporate object evolution support.

Our experience shows that exact duplicates only rarely occur, therefore the detection of inexact duplicates is of major importance. Currently, our rule-based approach only supports the detection of duplicates exactly matching a rule's definition, while our relation distance-based approach can be configured with a threshold that determines whether two objects are considered to be duplicates. By relaxing our current constraint that requires relations in a calculus to be pairwise disjoint and instead allowing multiple valid relations annotated with a probability to hold simultaneously, we plan to be able to not only state that, e. g., "an accident is partially overlapping with a traffic jam", but instead express that "an accident is partially overlapping with a traffic jam with a confidence of 60%, proper part of it with a confidence of 25%, and disrelated from it with a confidence of 15%". By this, we can partially match duplicate rules and give a confidence estimate on our duplicate detection. In addition, we plan to improve the distance-based approach's runtime behaviour as indicated above by incorporating clustering strategies. We will also investigate how this approach's scalability can be assured by evaluating how adding additional dimensions can enhance accuracy without hampering runtime behaviour.

The second major area we plan to address is the consideration of object evolution in duplicate detection. As our example in Sect. 2 shows, objects are not static and, therefore, similarity between them changes over time. Based on the knowledge on possible transitions in a CNGs and by regarding object evolution characteristics as described in [4], we aim to reconstruct object histories and reason on possible evolutions.

In addition, we plan to further evaluate the proposed duplicate detection method on real-life data from the Austrian highways agency ASFINAG[7] to verify our initial findings. For this, further issues identified in our survey [5] aside from the major areas pointed out above (e. g., data stream support allowing continuous duplicate detection

[7] www.asfinag.at

on live data in real time, explanations for human operators why objects are considered to be duplicates) need to be incorporated.

References

1. J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
2. C. Bailey-Kellogg and F. Zhao. Qualitative spatial reasoning - extracting and reasoning with spatial aggregates. *AI Magazine*, 24(4):47–60, 2003.
3. J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, 1983.
4. N. Baumgartner, W. Gottesheim, S. Mitsch, W. Retschitzegger, and W. Schwinger. On optimization of predictions in ontology-driven situation awareness. In *Proceedings of the 3rd International Conference on Knowledge, Science, Engineering and Management (KSEM)*, 2009.
5. N. Baumgartner, W. Gottesheim, S. Mitsch, W. Retschitzegger, and W. Schwinger. "Same, Same but Different"—A Survey on Duplicate Detection Methods for Situation Awareness. In *Proceedings of the 8th International Conference on Ontologies, DataBases and Applications of Semantics*, Vilamoura, Portugal, 2009. Springer.
6. N. Baumgartner, W. Gottesheim, S. Mitsch, W. Retschitzegger, and W. Schwinger. BeAware!—situation awareness, the ontology-driven way. *accepted for publication in: International Journal of Data and Knowledge Engineering*, 2010.
7. M. Bhatt, W. Rahayu, and G. Sterling. Qualitative simulation: Towards a situation calculus based unifying semantics for space, time and actions. In *Proceedings of the Conference on Spatial Information Theory*, Ellicottville, NY, USA, 2005.
8. J. Bleiholder and F. Naumann. Data fusion. *ACM Computing Surveys*, 41(1), 2008.
9. L. Chen, M. T. Özsu, and V. Oria. Robust and fast similarity search for moving object trajectories. In *Proceedings of the International Conference on Management of Data*, pages 491–502, New York, NY, USA, 2005. ACM.
10. B. L. Clarke. A calculus of individuals based on "connection". *Notre Dame Journal Formal Logic*, 22(3):204–218, 1981.
11. A. G. Cohn and J. Renz. *Handbook of Knowledge Representation*, chapter Qualitative Spatial Representation and Reasoning, pages 551–596. Elsevier Science Publishers Ltd., 2008.
12. Z. Cui, A. G. Cohn, and D. A. Randell. Qualitative simulation based on a logical formalism of space and time. In *Proceedings AAAI-92*, pages 679–684. AAAI Press, 1992.
13. M.-M. Deza and E. Deza. *Dictionary of Distances*. Elsevier Science Publishers Ltd., 2006.
14. F. Dylla and R. Moratz. Exploiting qualitative spatial neighborhoods in the situation calculus. In *Proceedings of Spatial Cognition IV Reasoning, Action, Interaction*, pages 304–322. Springer, 2005.
15. F. Dylla and J. O. Wallgrün. On generalizing orientation information in OPRA_m. In *Proceedings of the 29th Annual German Conference on AI (KI2006)*, Lecture Notes in Computer Science, pages 274–288, Bremen, Germany, 2007. Springer.
16. F. Dylla and J. O. Wallgrün. Qualitative spatial reasoning with conceptual neighborhoods for agent control. *Journal of Intelligent Robotics Systems*, 48(1):55–78, 2007.
17. M. Endsley. *Situation Awareness Analysis and Measurement*, chapter Theoretical Underpinnings of Situation Awareness: A Critical Review, pages 3–33. Lawrence Erlbaum Associates, New Jersey, USA, 2000.
18. C. Freksa. Conceptual neighborhood and its role in temporal and spatial reasoning. In M. Singh and L. Travé-Massuyès, editors, *Proceedings of the Imacs International Workshop on Decision Support Systems and Qualitative Reasoning*, pages 181–187, 1991.

19. C. Freksa. Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54(1):199–227, 1992.
20. E. Frentzos, K. Gratsias, and Y. Theodoridis. Index-based most similar trajectory search. In *Proc. of the 23rd Int. Conf. on Data Engineering*, pages 816–825. IEEE, 2007.
21. E. Frentzos, N. Pelekis, I. Ntoutsi, and Y. Theodoridis. *Mobility, Data Mining and Privacy—Geographic Knowledge Discovery*, chapter Trajectory Database Systems, pages 151–188. Springer, 2008.
22. P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. MIT Press, 2000.
23. J.-R. Hwang, H.-Y. Kang, and K.-J. Li. Searching for similar trajectories on road networks using spatio-temporal similarity. In *Proc. of the 10th East Euro. Conf. on Adv. in Databases and Inf. Sys.*, pages 282–295, Thessaloniki, Greece, 2006. Springer.
24. M. Johannesson. Combining integral and separable subspaces. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 447–452, Edinburgh, Scotland, UK, 2001. Lawrence Erlbaum Associates.
25. M. T. Keane, D. Hacket, and J. Davenport. Similarity processing depends on the similarities present: Effects of relational prominence in similarity and analogical processing. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Edinburgh, Scotland, UK, 2001. Lawrence Erlbaum Associates.
26. H. Kirschfink, J. Hernandez, and M. Boero. Intelligent traffic management models. In *Proccedings of the European Symposium on Intelligent Techniques (ESIT)*, pages 36–45, Aachen, Germany, September 2000.
27. M. M. Kokar, C. J. Matheusb, and K. Baclawski. Ontology-based situation awareness. *International Journal of Information Fusion*, 10(1):83–98, 2009.
28. G. Ligozat. Towards a general characterization of conceptual neighborhoods in temporal and spatial reasoning. In *Proceedings of the AAAI-94 Workshop on Spatial and Temporal Resoning*, pages 55–59, Seattle, WA, USA, 1994. AAAI.
29. M. Ragni and S. Wölfl. Temporalizing cardinal directions: From constraint satisfaction to planning. In *Proceedings of 10th International Conference on Principles of Knowledge Representation and Reasoning*, pages 472–480. AAAI Press, 2006.
30. E. Rahm and H. H. Do. Data Cleaning: Problems and Current Approaches. *IEEE Data Eng. Bull.*, 23(4):3–13, 2000.
31. Randell, D.A., Z. Cui, and A. G. Cohn. A spatial logic based on regions and connection. In *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning*. Morgan Kaufmann, 1992.
32. A. Schwering. Evaluation of a semantic similarity measure for natural language spatial relations. In *Proceedings of the 8th International Conference on Spatial Information Theory (COSIT)*, pages 116–132, Melbourne, Australia, 2007. Springer.
33. A. Schwering. Approaches to semantic similarity measurement for geo-spatial data: A survey. *Transactions in GIS*, 12(1):5–29, 2008.
34. N. van de Weghe and P. D. Maeyer. Conceptual neighborhood diagrams for representing moving objects. In *Proceedings of the ER Workshop on Perspectives in Conceptual Modeling*, pages 228–238, Klagenfurt, Austria, 2005. Springer.