Proceedings of the 1st International Workshop on

Context-Awareness in Geographic Information Services (CAGIS 2014)

in conjunction with GIScience 2014

Vienna, Austria

23 September 2014

Haosheng Huang, Jürgen Hahn, Christophe Claramunt, Tumasch Reichenbacher (Eds.)

Editors

Haosheng Huang, Jürgen Hahn
Department of Geodesy and Geoinformation
Vienna University of Technology
1040 Vienna
Austria
{haosheng.huang, juergen.hahn}@tuwien.ac.at

Christophe Claramunt
Naval Academy Research Institute
Lanvéoc-Poulmic,
BP 600, Brest naval
France
christophe.claramunt@ecole-navale.fr

Tumasch Reichenbacher
Geographic Information Visualization & Analysis (GIVA)
Department of Geography
University of Zurich - Irchel
Winterthurerstr. 190
CH-8057 Zurich
Switzerland
tumasch.reichenbacher@geo.uzh.ch

Preface

This document contains the proceedings of the 1st International Workshop on Context-awareness in Geographic Information Services (CAGIS 2014), held on September 23, 2014 in Vienna, Austria, in conjunction with the 8th International Conference on Geographic Information Science (GIScience 2014).

Recent years have seen an increasing interest in using geographic information services such as Location-Based Services (LBS) in assisting our daily behavior and decision-making. For effectively supporting users, these services should provide information and services adapted to a user's context, needs and preferences. For example, when recommending places for a tourist to visit, different context information such as weather, time (e.g., weekdays vs. weekends) and with whom (e.g., alone vs. with children vs. with others) is often needed to provide relevant results. From this sense, context-awareness and adaptation play a key role in geographic information services.

Questions related to understanding and dealing with the context in which a user is interacting with geographic information through an information service on a computer device (let it be a mobile phone, or a data center cluster, or everything in between) have been investigated by different lines of research within the field of Geographic Information Science, for example: Geographic Information Retrieval, Location-Based Services, Mobile Cartography, and Recommender Systems. Different approaches have been proposed in the literature to model and use context information in geographic information services.

This workshop aimed to provide a forum for these lines of research to meet and to discuss how context-awareness can be introduced into geographic information services to provide information and services that are adapted to a user's context. The objective was to outline an overview of the "state of the art" of context-awareness in geographic information services, as well as to identify and formulate key research questions for the future development of the field.

The CAGIS 2014 workshop featured two keynote talks, and eight presentations of papers. Each paper has been reviewed by three or more members of the Program Committee. We are grateful for the collaborative efforts of the authors, the members of the Program Committee, and our keynote speakers Carsten Keßler (Hunter College – CUNY) and Kazutoshi Sumiya (University of Hyogo). Finally, we would like to thank the GIScience organizers, especially Paolo Fogliaroni in helping organizing the workshop.

September 2014

Haosheng Huang
Jürgen Hahn
Christophe Claramunt
Tumasch Reichenbacher

Program Committee

Kate Beard, University of Maine, USA
Michela Bertolotto, University College Dublin, Ireland
Alexis Comber, University of Leicester, UK
Stefano De Sabbata, Oxford Internet Institute, UK
Paolo Fogliaroni, Vienna University of Technology, Austria
Christian Freksa, University of Bremen, Germany
Peter Fröhlich, Telecommunications Research Center Vienna, Austria
Georg Gartner, Vienna University of Technology, Austria
Sven Gehring, DFKI, Germany
Bo Huang, Chinese University of Hong Kong, Hong Kong
Ki-Joune Li, Pusan National University, South Korea
Xiang Li, East China Normal University, China
Farid Karimipour, University of Tehran, Iran
Carsten Kessler, Hunter College New York, USA
Miguel Félix Mata Rivera, National Polytechnic Institute, Mexico
Liqiu Meng, Technische Universität München, Germany
Mir Abolfazl Mostafavi, Laval University, Canada
Giuliana Pallotta, NATO STO CMRE, Italy
Cyril Ray, Naval Academy Research Institute, France
Tapani Sarjakoski, Finnish Geodetic Institute, Finland
Christoph Schlieder, University of Bamberg, Germany
Kazutoshi Sumiya, University of Hyogo, Japan
Taro Tezuka, University of Tsukuba, Japan
Bisheng Yang, Wuhan University, China

Organizing Committee

Chairs: Jürgen Hahn, Haosheng Huang, Vienna University of Technology, Austria

Christophe Claramunt, Naval Academy Research Institute, France
Tumasch Reichenbacher, University of Zurich, Switzerland

Table of Contents

Keynote Talks

Contributed Papers

Keynote talk 1

Research in the Age of the Context Machine

Carsten Keßler
Hunter College – CUNY, USA
carsten.kessler@hunter.cuny.edu

Abstract

One of the major challenges in the development of context-aware applications has always been the initial step of collecting enough information about a user's context. With the increasing prevalence of smartphones equipped with a plethora of sensors, more and more users have a context machine on them that constantly collects, uses, and transmits different kinds of passively collected contextual information. Additionally, many users actively provide contextual information by participating in online social networks. This talk will shed some light on the implications of these developments for research on context awareness. Starting with a brief review of the history of research in context awareness, it will discuss the role of research conducted in industry in this field, upcoming research challenges, and implications for user privacy.

Speaker's Bio

Carsten Keßler is an Assistant Professor for Geographic Information Science at the Department of Geography at Hunter College–CUNY in New York and Associate Director of the Center for Advanced Research of Spatial Information. He is one of the organizers of the Linked Science workshop series and co-chairing the W3C Emergency Information Community group. Before moving to New York in fall 2013, he was a post-doc researcher in the Semantic Interoperability Lab at Institute for Geoinformatics, University of Münster, Germany. His research interests are in the areas of context modeling, information integration, geospatial semantics, linked data, volunteered geographic information, emergency management, and participatory GIS.

Keynote talk 2

Less-Conscious Information Retrieval Techniques for Location Based Services

Kazutoshi Sumiya

University of Hyogo, Japan

sumiya@shse.u-hyogo.ac.jp

Abstract

We have developed methods which can deal with the users' interaction without conventional conscious searching. When a user generally performs map operations with certain information retrieval intentions (less-conscious), a system using our method can detect the specific operation sequences. For example, if the user performs zooming-in and centering operations, the user is narrowing down the search area to a certain location. We define such operation sequences as chunks. The system detects the chunks and uses them to analyze the user's operations and thereby detect the user's intentions. We have developed several prototype systems based on the proposed methods.

Speaker's Bio

Prof. Kazutoshi Sumiya received his BE and ME degrees in instrumentation engineering from Kobe University in 1986 and 1988, respectively. Then he joined Panasonic (Matsushita Electric Industrial Co). He received his Ph.D in Information media from Kobe University in 1998. He left the company and became a lecturer at Kobe University in 1999, and then was promoted to an associate professor in 2000. He became an associate professor in 2001 at Kyoto University and a professor at the University of Hyogo in 2004. At Kobe University and Kyoto University, he developed information dissemination systems and fusion technique for broadcast media and network media. At the University of Hyogo, he is developing next-generation information techniques. He is a chair of Database System special interest group (DBS) in the Information Processing Society of Japan (IPSJ) and a co-editor of IPSJ Transaction on Database.

A Place and Event Based Context Model for Environmental Monitoring

Kate Beard[1] and Melinda Neville[2]

[1]School of Computing and Information Science, University of Maine, Orono, Maine
[2]Ecology and Environmental Science, University of Maine, Orono, Maine
beard@spatial.maine.edu, mneville@maine.edu

Abstract. The importance of context awareness in support of computational services has been well recognized with applications in areas such as real-time location-based services, dynamic social network collaboration, situational health monitoring, indoor navigation, and the Internet of Things among others. The role of context in these services is generally to support more responsive service delivery for human users or agents. The focus of this paper is a context model for environmental observations, where knowledge of spatial and temporal contextual differences among observations is important for interpretation and analyses as well as facilitating sharing and reuse of data outside the original collection context. This paper builds on the OBOE ontology for observation data and expands spatial and temporal context settings through additional ontologies that support flexible spatial contextual consruction in terms of places and relationships among places and temporal contextual construction in terms of events and event relationships. The goal is to capture spatial and temporal contexts for observations to support machine as well as human interpretation and analysis.

Keywords: spatial-temporal context, ontology based context model, gazetteer

1 Introduction

Context awareness is employed to support intelligent decisions and automate responses to situations and events that have occurred. Context awareness is important for customizing information services in ways appropriate to users' characteristics, devices, spatial and temporal settings, and activities. Many context dependent services have been investigated including location based services, social media, driver assistance services, indoor navigation, and health care situation monitoring [1]. These applications have tended to focus on personalization and adaptability of services to user contexts. An area in which context is also very important but has received less attention is in the provision of context for scientific observations [2]. Environmental observations

Published in the Proceedings of the 1st International Workshop on
Context-Awareness in Geographic Information Services, in conjunction with
GIScience 2014, edited by Haosheng Huang, Jürgen Hahn, Christophe Claramunt,
Tumasch Reichenbacher, CAGIS 2014, 23 September 2014, Vienna, Austria, 2014.

are inevitably influenced by their spatial and temporal settings which could be more explicitly modeled for improved scientific interpretation and analysis of the data.

Contextual information for observation data is often only implicitly available through database schema labels, from plotting locations on a map, or through natural-language based metadata. Observation metadata may include location and time stamps as well as information about the observer and observation protocols. A GPS coordinate location is easy to collect and frequently serves as the reported location of an observation, but a coordinate alone offers little to no information about a spatial setting. Similarly a time stamp locates an observation in time but offers little in the way of temporal context for an observation. For example, water samples taken for water quality assessment may have different spatial settings such as near a point source, at the outlet of a particular stream, or in an urban versus a rural setting. Similarly, water samples have very different temporal context if taken before, during or after a precipitation event. While water quality experts are well aware of contextual related influences on water quality parameters, such context information is typically not explicitly captured in formats conducive for automated search and analysis.

What constitutes spatial and temporal context can be difficult to define and bound, and thus some flexibility and vagueness in specifying spatial and temporal context is desirable. This paper expands on previous work and the OBOE ontology for observation data [3] and explores the use of places, place to place relationships, events and event relationships as building blocks for an open ended spatial and temporal context model. Places are understood to refer to named instances of regions or features [4]. They need not have explicit spatial bounds or alternatively they may have many possible spatial representations. A gazetteer that models places and which is enhanced to model relationships among places is proposed as the basis for flexible spatial context model. Similarly, events and relationships among events provide the basis for temporal context development. Section 2 of the paper reviews previous work on concepts of context and context models. Section 3 presents the proposed place and event based spatial-temporal context model for observations. Section 4 illustrates the proposed observation context model for a specific environmental monitoring setting and Section 5 concludes with some issues for future research.

2 Review of Context and Context Models

Many definitions of context exist. Dey [5] defined context as any information used to characterize the situation of an entity where an entity could be a person, location, object, or event. Context in location based services, for example, typically identifies user location (an X,Y coordinate), an approximate neighborhood, who the user is with, and what activities they are engaged in. Context in wireless sensor networks (WSN) has been defined to include sensor node resources, network characteristics, network states, and energy management [6]. Context is also seen as having scale or different levels of detail encompassing local or fine to large and coarse scales or granularities [6]. One of the complexities of context is its own context dependencies, and a challenging issue is managing the dynamics of context which in a worst case may be in constant flux [7,8].

A context model has been described as a structure for the representation of situations in the real world for interpretation and exchange by machines. Context models specify the entities and relations among entities needed to characterize a situation or setting and have included representation as key-value pairs [5], object-role models [9], and spatial models [10,11,12].

Ontology based context models have been recently introduced to improve interoperability, reusability, and context based reasoning [9], [13, 14]. Several ontology based context models include similar context classes and entities. Becker and Nicklas [15] identify primary context as including identity of entities, location, and time, and most context models include these as high level classes [14], [16]. Becker and Nicklas [15] also characterize four primary ways in which context can be utilized. This paper focuses on context based tagging, the tagging of information to context to allow later action based on this context.

The development of context models has gone hand in hand with substantial growth in new sources of context information. New technologies, including smart phones, smart devices, and sensor networks serve as both consumers and providers of context information. A number of recent context models assume a sensor based information gathering layer where information from sensor streams is analyzed for recognized activities or events [16]. Data acquired from sensors is then used directly as low level contextual information or to reason and construct higher level contextual constructs through inference.

Ontology based context models typically specify a set of general context entities and relations [16], [18,19] and common to many of these is an explicit place or location class. The COMANTO context ontology [15] includes a Place class for representing an abstract or physical spatial region and also includes spatial relationships among places, such as adjacent to, included in, or a hierarchical place containment structure (e.g. cities containing buildings and streets). Temporal context has received a similar level of attention. A number of temporal models rely on events to represent conditions of interest or changes of state. Barreneachea et al [8] describe a distributed event-based system (DEBS) that employs loosely coupled components communicating via event-based asynchronous interactions. Andrienko et al [20] describe spatio-temporal context for movement data that links movement events through spatial and temporal relationships to other locations and events. Janowicz et al [21] emphasize the importance of space and time as contextual foundations in the Linked Data world.

Bowers et al [3] provide an ontology based model for scientific observation data including context specification. They define context as the meaningful surroundings of an observation, including other observations, their measured values, and their relationship to the observed entity. The OBOE ontology [2], illustrated in Figure 1, represents observations as assertions about entities, including one or more measurements, which assign a value to a characteristic of an entity. The OBOE ontology specifies context as a unary relation, hasContext, between observations. A context thus consists of named relationships between one observation and others indicating that an observation has been made within the scope of associated observations [3]. An example from [3] is a measure of diameter at breast height (DBH) as an observation on a tree. This observation is related by a "within" relation to a temporal observation measured in years and by another "within" relation to an observation of a plot as a location measured on a nominal scale.

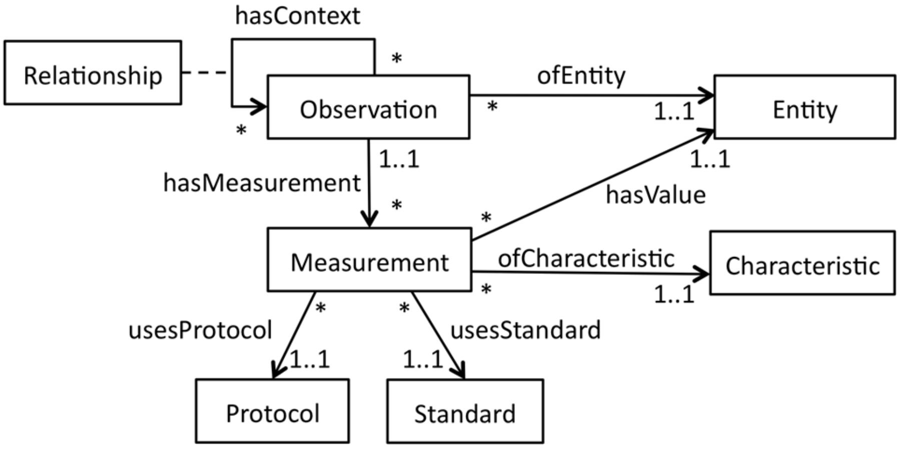


Fig. 1. The OBOE ontology for ecological observation data [3]

This is a flexible model in that any number and type of observation can be associated to create context. A limitation, however, is that no explicit classes of context are distinguished as in other context models [13], [15]. Specifically spatial and temporal context observations are not distinguished from any other context observations and are thus not searchable or retrievable as context components.

A second limitation of the OBOE context model is that it does not support higher level constructs. Context remains a Tier 2 level observable reality in Frank's [22] terms, rather than supporting higher level abstractions over observations. For example Bowers et al [3] give an air temperature observation as a context observation for a tree observation. A single air temperature observation or a daily average provides some level of context, but a sequence of observations abstracted as an event and indicating a period of rising or falling air temperature including the rate of rise or fall provides a higher level and richer context than a single observation.

3 Place and Event Based Context Model for Observation Data

The proposed context model builds on the OBOE and other context models with three objectives: 1) to make spatial and spatio-temporal context distinct (from other types of context), 2) to allow higher level constructs for creating context, and 3) allowing context to be indefinite, open ended, and context dependent. The approach reuses classes and relationships from OBOE including the Observation, Entity, Measurement, and Characteristic classes. In addition it utilizes the hasSpatialSetting and hasTemporalSetting relationships from the GEM model [23], the SpatialObject class from GeoSPARQL [24] and the TemporalEntity class from Owl-Time [25]. Semantic web technologies are used to implement the context model. An ontology based gazetteer implemented as an RDF triplestore supports spatial context construction and provides some level of reasoning over context information. The context model can be queried through SPARQL and GeoSPARQL.

3.1 The Spatial Context Model

The approach for making spatial context explicit and open-ended is managed by assigning an observation a spatial setting and then allowing the spatial setting to be expanded as appropriate. The assigned spatial setting is considered a local spatial context. An OBOE Observation is related to a GeoSPARQL SpatialObject with the hasSpatialSetting property from GEM as shown in Figure 2. The GeoSPARQL SpatialObject class has two subclasses: Feature and Geometry where Feature can be a distinct physical object in the landscape and may refer to a named geographic location or place [24].

Fig. 2. The OBOE observation class is connected through an OWL object property GEM:hasSpatialSetting to a GeoSPARQL SpatialObject class which can be a feature (place) or geometry.

The GeoSPARQL geometry class includes subclasses point, polyline, and polygon among others. The spatial setting for an observation can thus be a named place (feature), a geometry, or both. The GeoSPARQL ontology [24] defines an OWL object property, hasGeometry, between Feature and Geometry classes allowing a feature to be associated with zero or many geometries. A benefit of this approach is that a spatial setting need not imply any specific geometry allowing for some vagueness in the spatial setting. For example one might want to indicate that the setting of an observation is the mouth of a stream without having to specify such a setting with explicit geometry.

The assigned spatial setting can be extended along two possible pathways allowing for open ended context construction. If the assigned spatial setting is a geometry type it can be expanded through spatial relationships [26, 27] among geometry types as supported by GeoSPARQL. For example if the spatial setting of an observation is a polygon representing a field plot, a possible expanded spatial context could be the set of adjacent field plots. If the spatial setting is a place or feature type with no geometry and we wish to expand spatial context we need a mechanism to establish relationships between features. Here we focus on this second pathway as an important way to capture relationships among features or places not easily captured by spatial topological, distance, or directional relationships. This second expansion pathway relies on a semantically enhanced gazetteer that incorporates feature to feature and feature part-whole relationships. The enhanced gazetteer is developed from two ontologies: a geographic feature ontology and a gazetteer ontology. The approach has similarities to the SPIRIT project [28] which defined a three part ontology based model for geospatial search and in follow on work, [29] demonstrated expansion of place name search using spatial relationships such as near, north, south, east, or west of a place name.

The geographic feature ontology models prototypical features (places) and relationships among them. We illustrate the approach for a subdomain of hydrologic features modeled as subclasses of the GeoSPARQL Feature class. A class hierarchy of prototypical surface hydrology feature types is shown in Figure 3. FreshwaterBay and MarineBay are examples of prototypical feature parts. Namespace prefixes used in this example and elsewhere in the paper include geo: GeoSPARQL, hfo: HydrologicFeatureOntology, hgaz: HydrologicFeatureGazetteer.

Specification of feature to feature relationships is the important element that allows a feature or placed based spatial setting to be expanded to a broader spatial context. Specification of these as OWL properties makes use of OWL semantics to support context expansion through inference. The HFO includes OWL object properties; hasHydrologicRelation, hydrologicPartOf and its inverse, hasHydrologicPart, to express general associations between hydrologic feature classes. These general hydrologic relations are specialized by sub-properties (shown in Table 1) to express semantic feature-feature and feature-part relationships between the prototypical hydrologic feature classes. These feature-feature and feature-part relationships are instantiated in a hydrological feature gazetteer. These relationships are initially derived by GIS analysis but once instantiated in the gazetteer triple store they are easily accessed for context expansion without expensive spatial operations.

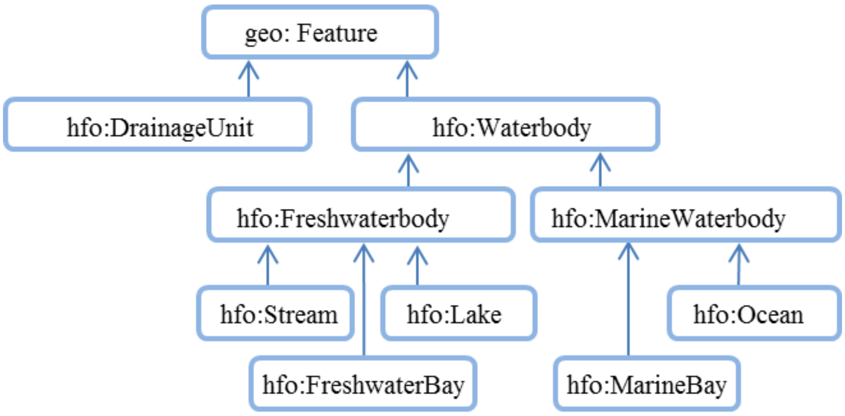


Fig. 3. Class hierarchy of prototypical hydrological feature types.

Table 1. Example OWL object properties, sub-properties and characteristics as specified in the HFO.

Property	SubProperty	Characteristics
hasHydrologicRelation	flowsInto	Antisymmetric, Intransitive
	flowsFrom	
	flowThrough	
	hasInflow	
	hasOutflow	
	isSourceOf	
	hasSource	
	hasMouth	
	isMouthOf	
	isTributaryOf	Antisymmetric, Transitive
	hasTributary	
hasHydrologicPart	hasFreshwaterBay	Antisymmetric, Transitive
	hasMarineBay	
hydrologicPartOf	FreshwaterBayOf	
	MarineBayOf	

Expansion of a place or feature based spatial setting is executed through queries to the gazetteer triplestore. Table 2 illustrates a SPARQL query template for expanding a feature based spatial setting. Through the SPARQL query, Maquoit Bay, an instance of a MarineBay is expanded to the set of feature instances hydrologically connected to Maquoit Bay. This set of feature instances and their relationships to the spatial setting feature form one possible spatial context. OWL semantics on these relationships such as the transitive property of hasTributary allows the relationships to be expanded to their transitive closure. Thus by inference a connected network of features can be obtained including tributaries of streams connected to Maquoit Bay, bodies of water they may flow through, and drainage units drained by the streams. The standard set of topological, directional, and proximity relationships would not as easily or directly obtain such a set of connected features and parts.

Table 2. SPARQL query template for constructing spatial context through feature to feature relations. The query starts from a named feature (e.g. Maquoit Bay) specified as the SpatialSetting and expands to semantically related features.

```
SELECT ?feature2 ?name2 ?hydrorel,
WHERE {
?feature1   hgaz: gnisname   "Maquoit Bay".
?feature1   rdf:type   ?fclass.
?hydrorel   rdfs:subPropertyOf   hfo:hasHydrographicRelation.  (gets specialized relationships)
?hydrorel   rdfs:range   ?fclass.             (gets relationships specific to Maquiot Bay's
feature class)
?feature2   ?hydrorel   ?feature1.            (gets the related features)
?feature2   hgaz:gnisName  name2}             (gets the related feature name)

SELECT ?hydrorel, ?feature2 ?name2
WHERE {
?feature1   hgaz: gnisname   "Maquoit Bay".
?feature1   rdf:type   ?fclass.
?hydrorel   rdfs:subPropertyOf   hfo:hasHydrographicRelation.  (gets specialized relationships)
?hydrorel   rdfs:domain  ?fclass.
?feature1   ?hydrorel   ?feature2.            (gets the features involved in the relationship)

?feature2   hgaz:gnisName ?name2}
```

3.2 Creating Temporal Context

Context in the OBOE model relies on relationships to individual observations and as a consequence misses important aspects of the temporal dimension. Parallel to the spatial context approach, we start with the specification of a temporal setting and allow this temporal setting to be expanded dynamically as needed by identifying events that are temporally related to the temporal setting.

A temporal setting is specified by connecting the OBOE observation class through the object property, hasTemporalSetting to the OWL-Time TemporalEntity class as illustrated in Figure 4. The OWL TemporalEntity has two subclasses, Instant and Interval, which allows a temporal setting to be either an instant or an interval.

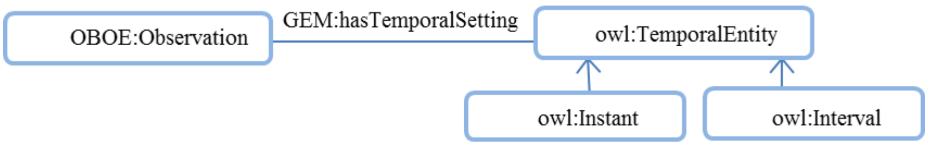


Fig. 4. An OBOE: observation class is connected through the GEM:hasTemporalSetting object property to the OWL-Time TemporalEntity class.

By specifying a temporal setting as a TemporalEntity, we make use of the semantics of OWL-Time. OWL-Time specifies a ProperInterval as a subclass of Interval. A ProperInterval is related to two Instants that specify a start time and end time through hasBeginning and hasEnd object properties. Temporal relationships as specified by Allen [30] can be asserted between ProperIntervals.

A number of context models rely on sensor data for context and context awareness and many rely on events abstracted from sensor data streams [14], [16]. For temporal context expansion we rely on an event database including events obtained as abstractions from sensor time series data. Such events are defined as subsequences of a sensor time series for which a particular property holds over a temporal interval [31]. For example, events extracted from a sensor time series of stream flow based on some domain defined threshold might include BaseFlow, HighFlow, and LowFlow events. Events from other sources, such as generated by human observation, or action (e.g. house construction), can also contribute to temporal context. Events are specified as a subclass of TemporalEntity which means they can be intervals or instants. Events are also assumed to have spatial settings, here specified by the GEM: hasSpatialSetting property to the GeoSPARQL SpatialObject class. Thus similar to an observation, an event can be situated in a place (feature) or by geometry (point, line, polygon). Events are also assumed to have some domain supported type classification.

Given an assigned temporal setting for an observation expressed as either an interval or instant, Temporal Context is the set of events in some temporal relation (e.g. before, concurrent) with an observation's temporal setting and additionally having relevant spatial relations to the observation's SpatialSetting or expanded spatial context. Standard SPARQL does not support temporal queries, but using extensions to SPARQL we can retrieve events of a specified type that have occurred within some temporal range of the observation's temporal setting. We can also retrieve events that have occurred in a specific temporal interval relationship (before, meet, overlap, during, start, finish, equal) to a temporal setting interval or in the case of an instant, before or equal [30].

Events that qualify for Temporal Context by satisfying temporal relationships must also be evaluated for relations with the observation's SpatialSetting and Spatial Context sets. Events are first checked to see if they share a spatial setting with an observation and then checked for spatial relationships with an observation's extended Spatial Context.

4 Context Model Example

To illustrate the spatial dimension of our context modeling approach we use water quality observations collected for shellfish harvest monitoring. Shellfish (clams, mussels, and oysters) are filter feeders, so the quality of the waters in which they grow is a key factor in determining whether they are safe to eat. Marine water samples are collected and tested throughout the year to evaluate levels of pathogenic bacteria and the presence of high levels trigger shellfish bed closures. This monitoring setting has a number of context dependencies important for understanding the spatial and temporal dynamics of coastal pollution events and bacterial outbreaks. Complex system interactions exist between natural process events such as precipitation, temperature, and salinity changes and anthropogenic events such as wastewater treatment protocols, sewer or storrmwater discharges, or changes in land use-land cover. Temporal dependencies can arise as a result of stream chemistry reacting differently to rainfall events depending on season and weather [32,33,34]. Spatial dependencies include catchment setting, size of embayment, and number and size of freshwater inputs.

The shellfish harvest area water quality observation data [35] include a station location, date, sampling protocol (e.g. R = random), a fecal coliform count per 100ml seawater sample (Score), a laboratory test method (MFCOL= membrane filtration), some related observations on tide level and wind direction, and an adversity factor that includes for example: P= rain or mixed precipitation anytime within past 2 days (i.e. thunderstorms, rainfall more than a drizzle); T= thawing snow and ice melt; S= sewage treatment plant malfunction or bypass events; W= Waterfowl (10 or more), domestic or wild animals (i.e., at the station or in close enough proximity to have a possible impact). Table 3 shows an example water quality observation record [35]. Some level of contextual information is provided by the adversity factor but not in any formal way.

Table 3. Example of a shellfish growing area water quality observation record.[35].

Station	Date	Strategy	Score	Method	Collector	Adversity	Wind	Tide
WJ001.50	5/3/2012	R	80	MFCOL	EXT	P	NE	HE

For demonstration purposes, a subset of these observations was semantically annotated using parts of the OBOE ontology with the addition of a spatial setting specification. A Hydrologic Feature Gazetteer instantiated with features from the US National Hydrography Database (NHD) provides the basis for spatial context expansion. Features in the gazetteer are uniquely identified by their Geographic Names Information System (GNIS) number and associated with an official GNIS name.

A water quality monitoring station was specified as a subclass of a GeoSPARQL feature and each observation was assigned a unique identifier based on its station number (e.g. WJ001.50.345678). Because observations are taken at stations and stations are fixed, the spatial setting of an observation has a two part specification; an observation is assigned a station as a spatial setting, and a station is then assigned one or more spatial settings that may be specified as places, geometry, or both as shown below.

```
oboe: observationWJ003.021345 gem: hasSpatialSetting geo:
StationWJ003.02
geo:StationWJ003.02 gem: hasSpatialSetting
hgaz:gnis570752 #specifies a feature
geo:StationWJ003.02 geo:hasGeometry geo: PointWJ003.02
```

Spatial Context for an observation can be obtained by expansion of the spatialSetting to semantically related features (places) or to spatially related geometries. To create an expanded SpatialContext for a water quality observation taken at StationWJ003.02 with the feature based SpatialSetting, Staples Cove, we use the SPARQL query template shown in Table 2 on the Hydrologic Feature Gazetteer. The query retrieves a set of features hydrologically related to Staples Cove. Staples Cove is the mouth of three streams, Frost Gully Brook, Concord Gully Brook and Kelsey Brook. These streams and their relationship to the spatial setting form one possible spatial context for this observation. Spatial Context could be further expanded through inference on defined relationships. This prototype supports an example set of feature types and relationships and as such provides one example of how a semantically enhanced gazetteer could be used for flexible spatial context construction.

5 Summary

To be most effectively used, scientific observations can benefit from spatial and temporal context models. Building on the OBOE semantics for observations [3], this paper describes a model for open ended spatial and temporal context building for observations. Defining what constitutes spatial or temporal context is a context dependent problem and placing exact bound on these can be limiting. We address the spatial aspect of the problem by allowing features or places and relationships among features and places to define spatial context from narrow settings to expanded settings as a function of feature-feature relationships. An example water quality data set collected as part of a shellfish monitoring program was used for proof of concept. Future research needs to develop the temporal context model more fully and test the approach on larger data sets and different contexts. SPARQL queries for interacting with the context model are cumbersome for an average user and could benefit from further research on graphical interfaces and query rewriting to facilitate construction and interaction with context sets. Further investigation of effective visualization of observations within spatial and temporal context sets would also benefit researchers in exploring their observation data.

References

1. Fook, V. F. S., Tay, S. C., Jayachandran, M., Biswas, J. and Zhang, D.: An ontology-based context model in monitoring and handling agitation behaviour for persons with dementia. In: Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM'06), Washington, DC (2006)
2. Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., Villa, F.: An ontology for describing and synthesizing ecological observation data. Ecol. Inform. 2, 279–296 (2007).
3. Bowers, S., Madin, J., Schildhauer, M.: A conceptual modeling framework for ex-pressing observational data semantics. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) Conceptual Modeling-ER. LNCS, vol. 5231, pp. 41-54. Springer, Heidelberg (2008)
4. Goodchild, M.F., Hill, L.L.: Introduction to digital gazetteer research. Int. J. Geogr. Inf. Sci. 22, 1039–1044 (2008).
5. Dey, A.K.: Understanding and Using Context. Pers. Ubiquitous Comput. 5, 4–7 (2001).
6. Liu, Y., Seet, B.-C., Al-Anbuky, A.: An Ontology-Based Context Model for Wireless Sensor Network (WSN) Management in the Internet of Things. J. Sens. Actuator Networks. 2, 653–674 (2013).
7. Chang, Y.-L., Barrenechea, E., Alencar, P.: Dynamic user-centric mobile context model. Digital Information Management (ICDIM), 2010 Fifth International Conference on. pp. 442–447. IEEE (2010).
8. Barrenechea, E.S., Alencar, P.S.C.: An Adaptive Context-Aware and Event-Based Framework Design Model. Procedia Comput. Sci. 5, 593–600 (2011).
9. Henricksen, K., Indulska, J.: Developing context-aware pervasive computing applications: Models and approach. Pervasive Mob. Comput. 2, 37–64 (2006).
10. Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., Riboni, D.: A survey of context modelling and reasoning techniques. Pervasive Mob. Comput. 6, 161–180 (2010).
11. Strang, T., Linnhoff-Popien, C.: Context Modeling Survey. In: Proceedings of the 6th International Conference on Ubiquitous Computing (UbiComp), Nottingham, UK (2004).
12. Singhal, A., Luo, J. and Zhu, W. "Probabilistic spatial context models for scene content understanding." Computer Vision and Pattern Recognition Proceedings. IEEE Computer Society Conference. Vol.1. IEEE. vol.1, no., pp.I-235, I-241. (2003).
13. Chen, H. Perich, F., Finin T. and Joshi, A.: SOUPA: Standard Ontology for Ubiquitous Pervasive Applications. In: Mobile and Ubiquitous Systems: Networking and Services, 2004. pp. 258-267. The First Annual International Conference on Mobile and Ubiquitous Systems, IEEE Boston, USA (2004).
14. Wang, X.H., Zhang, D.Q., Gu, T., Pung, H.K.: Ontology based context modeling and reasoning using OWL. Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on. pp. 18–22 (2004).
15. Becker, C. and Nicklas, D.: Where do spatial context-models end and where do ontologies start? A proposal of a combined approach. Workshop on Advanced Context Modelling, Reasoning and Management under the 6th Int.Conf. on Ubiquitous Computing (UbiComp 2004),Nottingham, UK (2004).
16. Roussaki, I., Strimpakou, M., Kalatzis, N., Anagnostou, M., Pils, C.: Hybrid context modeling: A location-based scheme using ontologies. Fourth Annu. IEEE Int. Conf. Pervasive Comput. Commun. Work. 2–7 (2006).

17. Tan, J., Zhang, D., Wang, X., Cheng, H.: Enhancing Semantic Spaces with Event-Driven Context Interpretation. In: Gellersen, H.-W., Want, R., and Schmidt, A. (eds.) Pervasive Computing SE - 6. pp. 80–97. Springer Berlin Heidelberg (2005).
18. Chen, H., Finin, T., Joshi, A.: An ontology for context-aware pervasive computing environments. Knowl. Eng. Rev. 18, 197–207 (2003).
19. Ejigu, D., Scuturici, M., Brunie, L., Liris-umr-cnrs, L., Lyon, I. De, Capelle, J.: An ontology-based approach to context modeling and reasoning in pervasive computing. Pervasive Computing and Communications Workshops, 2007. Fifth Annual IEEE International Conference on. pp. 14–19. , New York (2007).
20. Andrienko, G., Andrienko, N., Heurich, M.: An event-based conceptual model for context-aware movement analysis. Int. J. Geogr. Inf. Sci. 25, 1347–1370 (2011).
21. Janowicz, K., Bröring, A., Stasch, C., Schade, S., Everding, T., Llaves, A.: A RESTful proxy and data model for linked sensor data. Int. J. Digit. Earth. 6, 233–254 (2013).
22. Frank, A.: Ontology for Spatial Temporal Databases. In: M. Koubarakis et al (eds.) Spatio-temporal Databases – The CHOROCHRONOS approach. LNCS, vol. 2520, pp. 9-77, Springer Berlin Heidelberg (2003).
23. Worboys, M., Hornsby, K.: From Objects to Events: GEM, the Geospatial Event Model. In: Egenhofer, M., Freksa, C., and Miller, H. (eds.) Geographic Information Science 2004. LNCS, vol. 3234, pp. 327–343. Springer Berlin Heidelberg (2004).
24. Battle, R., Kolas, D.: Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. In: Janowics, K. (ed.) Semantic Web IOS Press, pp. 1–17 (2012).
25. Pan, F and Hobbs, J.: Temporal Aggregates in OWL-Time. In: Russell, I., and Markov, Z. (eds.)Proceedings of the 18th International Florida Artificial Intelligence Research Society Conference (FLAIRS), Clearwater Beach, Florida, pp. 560-565. AAAI Press (2005).
26. Egenhofer, M.J., Franzosa, R.D.: Point-set topological spatial relations. Int. J. Geogr. Inf. Syst. 5, 161–174 (1991).
27. Randell, D.A., Cui, Z., Cohn, A.G.: A Spatial Logic based on Regions and Connection. In: Kaufmann, M. (ed.) Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning. , Cambridge, MA (1992).
28. Jones, C.B., Abdelmoty, A., Finch, D., Fu, G., AND Vaid, S. The SPIRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing. In GIScience*:* Proceedings of Third International Conference. 2004, Lecture Notes in Computer Science. (Berlin:Springer) 125-139. (2004).
29. Fu, G., Jones, C.B. and Abdelmoty, A. Ontology-Based spatial query expansion in information retrieval. In *Proceedings of the 2005 OTM Confederated international conference on On the Move to Meaningful Internet Systems: CoopIS, COA, and ODBASE - Volume Part II* (OTM'05), Robert Meersman and Zahir Tari (Eds.), Vol. Part II. Springer-Verlag, Berlin, Heidelberg, 1466-1482. (2005.)
30. Allen, J.F.: Maintaining knowledge about temporal intervals. Commun. ACM. 26, 832–843 (1983).
31. Rude, A., Beard, K.: High-Level Event Detection. GIScience, LNCS, vol. 7478, pp. 160–172 (2012).
32. Murdoch, P. and Stoddard. J.: Chemical characteristics and temporal trends in eight streams of the Catskill Mountains, New York. Water, Air, Soil Pollut. 67:367-395 (1993).
33. Likens, G.E., Buso, D.C.: Variation in Streamwater Chemistry Throughout the Hubbard Brook Valley. Biogeochemistry. 78, 1–30 (2006).
34. Aubert, a. H., Tavenard, R., Emonet, R., de Lavenne, a., Malinowski, S., Guyet, T., Quiniou, R., Odobez, J.-M., Merot, P., Gascuel-Odoux, C.: Clustering flood events from

water quality time series using Latent Dirichlet Allocation model. Water Resour. Res. 49, 8187-8199 (2013).
35. Nutting, G.: Growing Area WJ Triennial Review. Maine Department of Marine Resources (2012).

Computing geographic relevance in mobile information services

Stefano De Sabbata and Tumasch Reichenbacher

Department of Geography, University of Zurich, Switzerland
stefano.desabbata@geo.uzh.ch
tumasch.reichenbacher@geo.uzh.ch

Abstract. This paper presents a detailed account of a computational model for the assessment of geographic relevance, which is based on our previous work on a conceptual model and criteria of geographic relevance. The computational model implements five criteria, among those we empirically validated in an earlier publication, and it has been tested through a "user-centred" evaluation procedure presented – which is not presented in the current paper. Our aim is to foster the discussion on how to compute and combine scores to better assess the geographic component of situational relevance concepts in mobile information services.

1 Introduction

The aim of this paper is to provide a detailed description of a model for the quantitative assessment of geograhic relevance (GR) ([10, 12, 13]) as a combination of five among the criteria of GR, whose selection is based on the empirical results presented in [2]. An empirical validation of the model is discussed in [1]. The described model is focused on spaital and temporal components of relevance ranking, aiming to provide better information to users of mobile information service in time-constrained situations, while avoiding user profiling.

Studies support that the criteria topicality and spatio-temporal proximity are the most fundamental criteria of GR, and are hence included in the computational model. De Sabbata and Reichenbacher [2] suggest the criterion directionality, which is also included. At the same time, the criterion directionality is not completely independent from the criterion spatio-temporal proximity, as both criteria take into account the user's destination.

In addition to these first three criteria, we include the criteria cluster and co-location in this assessment model of GR, at the expense of more traditional IR criteria. There are three compelling reasons for this. First, the results presented in [2] show how cluster and co-location are among the primary criteria of GR. Second, we consider it important to deepen the understanding of criteria which have never been tested before. Third, these criteria convey information about the geographic environment of the entities, which is a distinguishing component of GR. A step in this direction has been undertaken by [7], who implement the five criteria as proposed in [2] (i.e., hierarchy, cluster, co-location, association-rule,

Published in the Proceedings of the 1st International Workshop on
Context-Awareness in Geographic Information Services, in conjunction with
GIScience 2014, edited by Haosheng Huang, Jürgen Hahn, Christophe Claramunt,
Tumasch Reichenbacher, CAGIS 2014, 23 September 2014, Vienna, Austria, 2014.

and anchor-point proximity) as binary filters within the "Cinemappy" mobile application for movie recommendations.

The computational model is based on three assumptions. First, for each criterion, it is possible to compute a 'distance value', which estimates the distance or difference between the user's information need and an entity under relevance assessment, with respect to the criterion under consideration. A semantic distance is taken into account for the criterion topicality. Spatial and temporal distances are taken into account for the criteria spatio-temporal proximity and directionality. Spatial distances and numerical differences (e.g., difference in the cardinality of two sets of objects) are taken into account for the criteria cluster and co-location. Second, based on the 'distance value' computed for a given criterion, it is possible to compute a 'score' (i.e., a numerical value normalised in the interval $[0, 1]$), which estimates the strength of GR, with respect to the criterion under consideration. This understanding of GR is inspired by the first law of geography [14], which is extended from the conventional geographic space to the non-geographic space of each criterion. For instance, concepts which are near in the semantic space (i.e., concepts whose semantic distance is small) are assumed to be more related than concepts that are distant in the semantic space (i.e., concepts whose semantic distance is large). Hence, the lower the 'distance value' computed for a criterion, the higher the 'score'. Third, a numerical estimation of GR can be computed combining the 'scores' described above. These

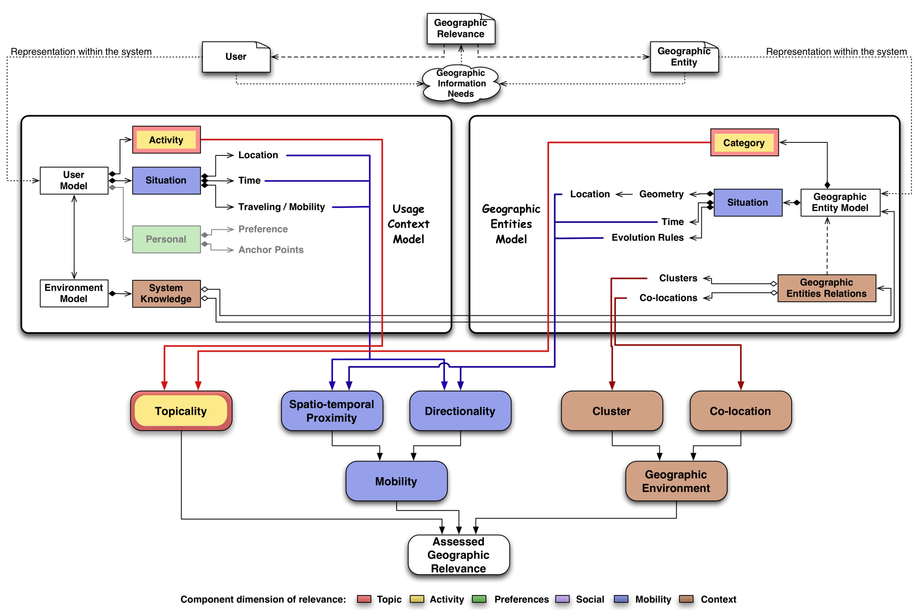


Fig. 1. Computational framework for the assessment of GR.

assumptions are taken into account within the scope of assessing GR as defined in [13].

Therefore, GR is assessed through a numerical estimation and combination of the criteria topicality, spatio-temporal proximity, directionality, cluster, and co-location. In turn, the numerical estimation of each criterion is grounded on the conceptual model presented by Reichenbacher *et al.* [12], as illustrated in Figure 1. The top half of the illustration reports a reduced version of the conceptual model presented by Reichenbacher *et al.* [12]. The bottom part in Figure 1 depicts how GR is derived from the elements of the conceptual model by means of the selected criteria.

Topicality takes into account a user's activity, and the category an entity belongs to, in order to achieve an estimated score for the topic and activity components of GR. Spatio-temporal proximity and directionality account for the situational elements in the user model and geographic entities model, in order to derive the respective estimated scores. These two scores are combined into an estimated score for the mobility component of GR. Cluster and co-location account for the relationships among geographic entities, which are the part of the environment model included in the geographic entities model, in order to derive the respective numerical scores. These last are combined in an estimated score related to the geographic environment of entities, that is part of the context component of GR. Finally, the scores computed for each component are aggregated into a final GR score, according to their relative importance, as discussed in [2].

Sections 2 and 3 issue formal definitions of the numerical scores associated with the selected criteria. Finally, Section 4 descusses various methods to combine the numerical scores defined in the previous sections in an aggregate value of geograhic relevance.

2 Base definitions

The definitions presented below are introduced in order to avoid long and ambiguous descriptions of the functions used, although a fully grounded mathematical definition of the problem is out of the scope of this paper.

$$
\begin{aligned}
q \in Q &: \text{ user query} \\
G = \{g_1, g_2, \dots\} &: \text{ geographic entities} \\
cat(g) &: \text{ category which } g \text{ belongs to} \\
dist(g_1, g_2) &: \text{ spatial distance between entities } g_1 \text{ and } g_2 \\
\Phi = \{\phi_1, \phi_2, \dots\} &: \text{ clusters of geographic entities} \\
\phi(g) &: \text{ cluster which } g \text{ belongs to} \\
\Phi^{cat(g)} = \{\phi_i, \phi_{i+1}, \dots\} &: \text{ clusters of geographic entities belonging to} \\
&\quad \text{the same category as a given entity } g
\end{aligned}
$$

$$\begin{aligned}
\Psi = \{\psi_1, \psi_2, \dots\}: &\ \text{co-location rules regarding the geographic entities} \\
&\ \text{where } \psi^p \text{ is the "premise" of a rule } \psi \\
&\ \text{and } \psi^c \text{ is the "conclusion" of a rule } \psi \\
\Psi(g) = \{\psi_j, \psi_{j+1}, \dots\}: &\ \text{co-location rules having as "premise"} \\
&\ \text{the same category to which } g \text{ belongs to} \\
t_{Clust}: &\ \text{threshold used for mining the clusters} \\
t_{Coloc}: &\ \text{threshold used for mining the co-location rules}
\end{aligned}$$

The user query element includes the user's current location, a destination, the activity that needs informational support, including the minimum time needed to perform this activity. The geographic entity elements contain all available information about the entities, including their category, location, and temporal availability. The remaining elements refer to the information obtained mining the dataset for spatial clusters and co-location rules (see [2]).

As from the assumptions advanced in Section 1, for each of the selected criteria, it is possible to compute a "distance function" δ for each criterion, such as:

$$\delta_{Criterion} : Q \times G \rightarrow \mathbb{R}_0^+ \tag{1}$$

which takes a user query and a geographic entity as input. The output value grows as the relevance of the geographic entity for the user context declines, in the scope of the considered criterion. Furthermore, for each of the selected criteria, it is possible to compute a "normalised score" such as:

$$\overline{s}_{Criterion} = f \circ \delta : Q \times G \rightarrow [0 \dots 1] \tag{2}$$

which is a function of the "distance function" defined above. Such a function takes a user query and a geographic entity as input, and returns a value between 0 and 1. The value 1 is assigned to the most relevant geographic entity for the user query in the scope of the considered criterion, and the value 0 is assigned to geographic entities which are completely irrelevant for the user query in the scope of the considered criterion.

3 Criteria scores

3.1 Topicality

A semantic distance function will be taken into account as the basis for computing the criterion topicality.. A survey of the literature on methods for measuring semantic relatedness is presented by [17]. In the scope of the current definition, the existence of such a semantic similarity distance function $\sigma(q, g)$ will be assumed as given.

The function $\delta_{Topicality}$ and $d_{Topicality}$ take into account the semantic distance to calculate the strength of the relationship between a user query and a geographic entity for the criterion topicality as shown in Equations 3 and 4. These formulas use the function $y = e^{(-\lambda \cdot x)}$ to transform a distance in the range $[0, +\infty]$ into a similarity value in the range $[0, 1]$, where λ regulates the steepness of the decrease. The score for the criterion topicality $\overline{s}_{Topicality}$ is defined as reported in Equation 5.

$$\delta_{Topicality}(q, g) = 1.0 - e^{(-\lambda \cdot \sigma(q,g))} \tag{3}$$

$$d_{Topicality}(q, g) = e^{(-\lambda \cdot \sigma(q,g))} \tag{4}$$

$$\overline{s}_{Topicality}(q, g) = \frac{d_{Topicality}(q, g)}{max_{j \in G}(d_{Topicality}(q, j))} \tag{5}$$

3.2 Spatio-temporal proximity

The criterion spatio-temporal proximity estimates the proximity between a geographic entity and the user in space and time. Assuming that the user is at a given location, and willing to be at a given destination by a given time, then a perfect match would be at any location along their travel trajectory.

The chosen approach is to compare the amount of time the user needs to perform their activity with the amount of time that they will be able to spend at the location of the entity, considering the travel time needed to reach it and be able to arrive at the destination on time, considering the temporal availability of the entity. The calculations are based on the space-time prism concept [6], as already suggested in the LBS field by [11]. A similar approach has also been adopted by [8].

The δ_{STprox} distance function takes into account a user's position, the location of an entity, a defined travel speed, a user schedule (i.e., a destination with a mandatory arrival time), a defined minimum time needed to accomplish the activity, and the time validity of the entity (e.g., opening hours). The distance function is then calculated as the ratio between the time needed to fulfil the activity, and the time a user is able to spend at the location of an entity, while the entity is also available (see Equation 6).

The following assumption is also considered: utility grows less than linearly, as the distance value decreases. That is, if an entity is available for twice as long as the user needs, the distance value is cut by half, but the entity is not twice as useful. Thus, the auxiliary function d_{STprox} is defined as a square root function of the inverse of the distance (see Equation 7). If the entity is available for less than the time specified by the user as necessary to perform the activity, the utility is zero. The score for the criterion spatio-temporal proximity is defined as shown in Equation 8.

$$\delta_{STprox}(q, g) = \frac{time\ needed}{time\ available} \tag{6}$$

$$d_{STprox}(q,g) = \begin{cases} 0.0 & \text{if } \delta_{STprox} > 1 \\ \sqrt{\dfrac{1}{\delta_{STprox}}} & \text{otherwise} \end{cases} \tag{7}$$

$$\overline{s}_{STprox}(q,g) = \frac{d_{STprox}(q,g)}{max_{j \in G}(d_{STprox}(q,j))} \tag{8}$$

It should be noted that the functions proposed in Equations 6, 7, and 8 could produce undesirable results when implemented in a real-world application. For instance, temporally unavailable entities would cause division by zero in Equation 6, and entities with non-finite temporal availability (e.g., shops open 24/7) would cause division by infinite in Equation 6, or division by zero in Equation 7. A specific handling of such cases would be necessary for the implementation of a reliable real-world application.

3.3 Directionality

The criterion directionality is implemented as a function of the angle between a user's destination and a geographic entity. That is, the angle between a straight line connecting a user's locationand a destination, and a straight line connecting the user's location and the location of the entity. The smaller the angle, the lower the value of the distance function δ_{AngDev} (see Equation 9), and the higher the value of the auxiliary function d_{AngDev} (see Equation 10). The function $cos(\alpha)$ is used in order to obtain a value equal to 1 when the angle α is 0, whereas the distance function δ_{AngDev} returns a value equal to 0 when the angle α is 0 (assuming $0 \leq \alpha \leq 180$). The resulting value is lower than 0.5 when the angular deviation is higher than 90. The score for the criterion directionality is calculated as reported in Equation 11.

$$\delta_{AngDev}(q,g) = \frac{1 - cos(\alpha)}{2} \tag{9}$$

$$d_{AngDev}(q,g) = \frac{cos(\alpha) + 1}{2} \tag{10}$$

$$\overline{s}_{Direct}(q,g) = \frac{d_{AngDev}(q,g)}{max_{j \in G} d_{AngDev}(q,j)} \tag{11}$$

3.4 Cluster

Two complementary aspects will be taken into account to implement the criterion cluster. The cardinality of the cluster (i.e., the larger the cluster, the higher the score) and the distance between the entity and the closest other entity of the same category (the shorter the distance, the higher the score). The combination of these two aspects conveys information on both the size and the density of the cluster. Moreover, the second aspect entails information about the relationship between an entity and other entities of the same category.

Assuming that the clusters have been mined for the dataset under investigation, the distance function $\delta_{ClustCard}(q, g)$ (see Equation 12) is calculated as the ratio between the number of entities in the largest among the clusters $\Phi^{cat(g)}$ of entities belonging to the same category as the entity under assessment g, minus the number of entities in the cluster $\phi(g)$ containing g, and the number of entities in the largest among the clusters $\Phi^{cat(g)}$ of entities belonging to the same category as g. It should be noted that in a real-world application the result of this distance function is influenced by the selection of the area taken into account during the cluster mining process. The distance function $\delta_{ClustDist}(q, g)$ (see Equation 13) is calculated as the ratio between the distance between g and the closest entity of the same category, and the distance t_{Clust} used as threshold for mining the clusters. This creates a relationship between the two aspects of the criterion mentioned above. Two auxiliary functions are then calculated as shown in Equations 14 and 15.

The score for the criterion cluster is calculated as a geometric combination of the two values calculated with the auxiliary functions (see Equations 16 and 17), as the distance on the Cartesian plane between the origin and the point described by the two values (see Equation 16). This approach is a compensatory version of the method used within the SPIRIT Project [16, 9], and thus it allows a disjunctive combination of the two aspects, which is necessary to obtain non-zero scores for those entities that are not in a cluster, as mentioned above.

There are some categories of entities which tend not to cluster, but instead are almost equally distributed over geographic space (e.g., pharmacies). In such cases, if no cluster has been identified for a given category, all entities of that category will obtain score 1 for the criterion cluster.

$$\delta_{ClustCard}(q, g) = \frac{max(\{\|\phi\| \mid \phi \in \Phi^{cat(g)}\}) - \|\phi(g)\|}{max(\{\|\phi\| \mid \phi \in \Phi^{cat(g)}\})} \tag{12}$$

$$\delta_{ClustDist}(q, g) = \frac{min(\{dist(g, h) \mid cat(g) = cat(h)\})}{t_{Clust}} \tag{13}$$

$$d_{ClustDist}(q, g) = e^{(-\lambda \cdot \delta_{ClustDist})} \tag{14}$$

$$d_{ClustCard}(q, g) = 1 - \delta_{ClustCard} \tag{15}$$

$$f_{Clust}(q, g) = \frac{\sqrt{d_{ClustDist}(q, g)^2 + d_{ClustCard}(q, g)^2}}{\sqrt{2}} \tag{16}$$

$$\overline{s}_{Clust}(q, g) = \begin{cases} 1.0 & \text{if } \Phi^{cat(g)} = \emptyset \\ 0.0 & \text{if } \Phi^{cat(g)} \neq \emptyset \wedge \delta_{ClustCard} = 1 \\ f_{Clust}(q, g) & \text{otherwise} \end{cases} \tag{17}$$

3.5 Co-location

A co-location rule is composed by a "premises" category and a "conclusion" category. Each rule captures the fact that, given an entity belonging to the first category, it is probable to find an entity belonging to the second category within a pre-defined distance – which has been used as threshold in the mining process.

We assume that a set of co-location rules have been identified, such that the affordances related to the "conclusion" category can be considered as related to (i.e., subsidiary, complementary, or consequent to) the affordances related to the "premises" category. This would imply a second-order relationship of relevance between the user's query and the entities belonging to the "conclusion" category.

The definition of the criterion co-location is similar to the one given for the cluster criterion in Section 3.4. Given a geographic entity under assessment, the criterion co-location selects the rules in which the category of that entity appears as "premises". For each of these rules, two aspects are considered. The first one is the number of entities within a threshold distance from the entity under assessment, that belong to the "conclusion" category of the rule. The second one is the distance between the entity under assessment, and the closest entity belonging to the "conclusion" category of the rule.

Thus, the distance function $\delta_{Coloc^{\psi}Dist}(q, g)$ (see Equation 20) for the rule $\psi \in \Psi$ is defined as the ratio between the distance between the entity g and the closest entity belonging to the "conclusion" category ψ^c of the rule ψ, and the distance threshold t_{Coloc} used in the mining process. The maximum cardinality of ψ (see Equations 18 and 19) is defined as the maximum number of entities belonging to the "conclusion" category ψ^c, within the threshold t_{Coloc} used in the mining process from an entity belonging to the "premises" category ψ^p. The distance function $\delta_{Coloc^{\psi}Card}(q, g)$ (see Equation 21) for the rule ψ is defined as the ratio between the difference between the maximum cardinality of ψ, and the number of entities belonging to the "conclusion" category ψ^c within the threshold t_{Coloc} used in the mining process, and the maximum cardinality of ψ. It should be noted that in a real-world application, the result of this distance function is influenced by the choice of the area taken into account during the co-location rules mining process.

Two auxiliary functions are calculated as shown in Equations 22 and 23. The definition of the distance-related similarity function $d_{Coloc^{\psi}Dist}(q, g)$ uses the same exponential function used for calculating topicality in Equation 3 and the first cluster score in Equation 14. The core of the auxiliary function $d_{Coloc^{\psi}Card}(q, g)$ is an inverse function of the value resulting from $\delta_{Coloc^{\psi}Card}(q, g) \in [0, 1]$, which is then decreased by 0.5 and finally multiplied by 2 in order to obtain 1 if $\delta_{Coloc^{\psi}Card}(q, g) = 0$ and 0 if $\delta_{Coloc^{\psi}Card}(q, g) = 1$.

For each of the co-location rules taken into account in assessing the GR of an entity, a further function $f_{Coloc^{\psi}}(q, g)$ is defined as a disjunctive geometric combination of the values obtained from the auxiliary functions defined above (see Equation 24). Finally, the co-location score for the entity under assessment is calculated as the average of the scores related to the different rules, as shown in Equation 25. As for the criterion cluster, if no rule has been identified which involve the entity's category as premises, then all the entities of that category are assigned the score 1.

$$card(\psi, x) = \|\{y \mid dist(x, y) \leq t_{Coloc} \wedge cat(x) = \psi^p \wedge cat(y) = \psi^c\}\| \quad (18)$$

$$maxCard(\psi) = max(\{card(\psi, x) \mid x \in G \wedge cat(x) = \psi^p\}) \quad (19)$$

$$\delta_{Coloc^\psi Dist}(q, g) = \frac{min(\{dist(g, h) \mid cat(h) = \psi^c\})}{t_{Coloc}} \quad (20)$$

$$\delta_{Coloc^\psi Card}(q, g) = \frac{maxCard(\psi) - \|\{h \mid dist(g, h) \leq t_{Coloc} \wedge cat(h) = \psi^c\}\|}{maxCard(\psi)} \quad (21)$$

$$d_{Coloc^\psi Dist}(q, g) = e^{(-\lambda \cdot \delta_{Coloc^\psi Dist})} \quad (22)$$

$$d_{Coloc^\psi Card}(q, g) = \left(\frac{1}{1 + \delta_{Coloc^\psi Card}} - \frac{1}{2}\right) * 2 \quad (23)$$

$$f_{Coloc^\psi}(q, g) = \frac{\sqrt{d_{Coloc^\psi Dist}(q, g)^2 + d_{Coloc^\psi Card}(q, g)^2}}{\sqrt{2}} \quad (24)$$

$$\overline{s}_{Coloc}(q, g) = \begin{cases} 1.0 & \text{if } \Psi^{cat(g)} = \emptyset \\ \sum_{\psi \in \Psi(g)} \frac{1}{\|\Psi(g)\|} \cdot f_{Coloc^\psi}(q, g) & \text{otherwise} \end{cases} \quad (25)$$

4 Scores' Combination

Assuming that the user is searching for an entity that satisfies all the selected criteria as well as possible, the simplest approach is to combine the normalised scores using the arithmetic product to combine the scores. Unfortunately, this method has a non-compensatory nature; that is, one low score is sufficient to obtain a low aggregate score. This would not be appropriate to combine the geographic environment component (cluster and co-location) with topicality and the mobility component (spatio-temporal proximity and directionality). This combination could produce false irrelevant cases, as the strong "and-ness" of the combination would cause possibly relevant entities to be scored as absolutely irrelevant. Moreover, this method causes a nonlinear distortion, which is undesirable in most cases.

Assuming a disjunctive approach to combining the scores, the arithmetic sum could be used instead. The geometric combination method used in the SPIRIT Project [16, 9] could also be adapted to account for more than two scores to combine. These methods have a compensatory nature; that is, one high score is sufficient to obtain a medium or high aggregate score. Although this behaviour can be appropriate when all the scores to be combined have the same importance,

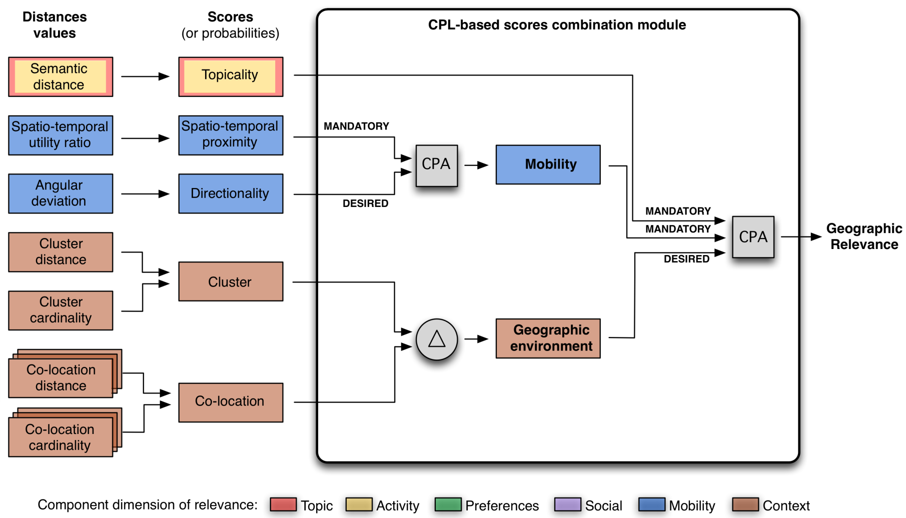


Fig. 2. Illustration of the GR assessment method based on the CPL model and GCD functions.

this is not the case for GR, where the topicality and the mobility components are more important than the geographic environment component.

In order to overcome these issues, the Continuous Preference Logic (CPL) model [3, 4] is used. This is a continuous logic of decision models, based on the generalised conjunction/disjunction (GCD) function [5]. Logic operators with any grade of partial conjunction $\triangle_\alpha$ (i.e. "and-ness") and partial disjunction $\triangledown_\omega$ (i.e., "or-ness") can be created using the GCD function.

CPL builds on such operators to define the conjunctive partial absorption (CPA) and the disjunctive partial absorption (DPA) operators. The CPA operator allows to combine "mandatory" input with "desired" input in a conjunctive manner (see Equation 26). The input defined as "mandatory" is accounted as starting value, which is then incremented or decremented, depending on whether the input defined as "desired" is greater or lower than the "mandatory" input, and on the "and-ness" of the partial conjunction and the "or-ness" of the partial disjunction used. If the "mandatory" input is zero, the output will always be zero. Similarly, the DPA operator allows to combine "sufficient" input with "desired" input in a disjunctive manner (see Equation 27). The input defined as "sufficient" is accounted as starting value, which is then incremented or decremented, depending on whether the input defined as "desired" is greater or less than the "sufficient" input, and on the "or-ness" of the partial disjunction and the "and-ness" of the partial conjunction used. If the "sufficient" input is equal to 1, the output is always equal to 1. More complex operators can be created combining more GCD functions and the CPA and DPA operators (see [4]).

$$CPA_{\alpha\omega}(x_{\text{mandatory}}, y_{\text{desired}}) = x_{\text{mandatory}} \bigtriangleup_{\alpha} (x_{\text{mandatory}} \bigtriangledown_{\omega} y_{\text{desired}}) \quad (26)$$

$$DPA_{\omega\alpha}(x_{\text{sufficient}}, y_{\text{desired}}) = x_{\text{sufficient}} \bigtriangledown_{\omega} (x_{\text{sufficient}} \bigtriangleup_{\alpha} y_{\text{desired}}) \quad (27)$$

We define the score combination for the assessment of GR as illustrated in Figure 2. This definition is based on the importance of the criteria discussed in [2], and takes advantage of the CPL model and GCD functions. Starting from the left-most side, the "distances" calculated by means of the δ functions defined in Section 3 are taken as input to calculate the normalised scores $\overline{s}$. The scores are thereafter taken as input by the CPL-based scores combination module. The mobility component is calculated using a CPA operator, taking into account spatio-temporal proximity as "mandatory" input, and directionality as "desired" input (see Equation 28). The geographic environment component is calculated as a partial conjunction of cluster and co-location (see Equation 29). The resulting value will lie between the minimum and the average of the two input values, depending on the chosen "and-ness". Finally, a CPA operator takes into account topicality and the mobility component as "mandatory" input, including the geographic environment as "desired" input (see Equation 30), to return the estimated GR as an output score (see right-most side of the Figure 2).

$$\overline{s}_{Mobility}(q, g) = CPA_{0.75\,0.75}(\ \overline{s}_{STprox}(q, g),\ \overline{s}_{Direct}(q, g)\ ) \quad (28)$$

$$\overline{s}_{GeoEnv}(q, g) = \overline{s}_{Clust}(q, g)\ \bigtriangleup_{0.75}\ \overline{s}_{Coloc}(q, g) \quad (29)$$

$$GR(q, g) = CPA_{0.75\,0.75}(\ \{\ \overline{s}_{Topicality}(q, g),\ \overline{s}_{Mobility}(q, g)\ \},\ \overline{s}_{GeoEnv}(q, g)\ ) \quad (30)$$

5 Discussion

In this paper we presented a GR assessment model, which is derived from the conceptual model presented in [12], and entails the criteria topicality, spatio-temporal proximity, directionality, cluster, and co-location. We identified this as the minimal set including the fundamental and distinguishing criteria of GR, based on the results presented in [2]. We issued a formal definition of the scores related to each criterion, and proposing a schema to combine them into an aggregate estimation of GR, based on the CPL model and GCD functions.

In order to evaluate the proposed approach, the GR assessment model was prototypically implemented, including the algorithms used to mine clusters and co-location rules. In [1], the authors investigate the effectiveness of the proposed GR assessment method against two baseline methods, that resamble common simple approaches used in Location Based Services, following a crowdsourcing approach as described in [15]. A pooling system was used to select the geographic entities to be used in the three scenarios prepared for a "user-centred" evaluation procedure. The crowdsourced data were used to produce a "ground truth" rank, which was compared with the three methods used in the pooling phase. The results support the effectiveness of the GR assessment method proposed in this paper, and the inadequacy of the other three tested methods.

References

1. S. De Sabbata. *Assessing Geographic Relevance for Mobile Information Services*. PhD thesis, University of Zurich, 2013.
2. S. De Sabbata and T. Reichenbacher. Criteria of geographic relevance: an experimental study. *International Journal of Geographical Information Science*, 26(8):1495–1520, 2012.
3. JJ Dujmovic. Extended continuous logic and the theory of complex criteria. *J. University of Belgrade, Series on Mathematics and Physics*, 537:197–216, 1975.
4. J.J. Dujmovic. Continuous preference logic for system evaluation. *Fuzzy Systems, IEEE Transactions on*, 15(6):1082–1099, 2007.
5. J.J. Dujmovic and H.L. Larsen. Generalized conjunction/disjunction. *International Journal of Approximate Reasoning*, 46(3):423–446, 2007.
6. H.J. Miller and S.A. Bridwell. A field-based theory for time geography. *Annals of the Association of American Geographers*, 99(1):49–75, 2009.
7. V. C. Ostuni, T. Di Noia, R. Mirizzi, D. Romito, and E. Di Sciascio. Cinemappy: a Context-aware Mobile App for Movie Recommendations boosted by DBpedia. In *1st International Workshop on Semantic Technologies meet Recommender Systems & Big Data (SeRSy 2012)*. CEUR-WS, 2012.
8. P. Pombinho, M.B. Carmo, and A.P. Afonso. Context aware point of interest adaptive recommendation. In *Proceedings of the 2nd Workshop on Context-awareness in Retrieval and Recommendation*, pages 30–33. ACM, 2012.
9. R. Purves, P. Clough, C.B. Jones, A. Arampatzis, B. Bucher, D. Finch, G. Fu, H. Joho, A.K. Syed, S. Vaid, et al. The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. *International Journal of Geographical Information Science*, 21(7):717–745, 2007.
10. J. Raper. Geographic relevance. *Journal of Documentation*, 63(6):836–852, 2007.
11. M. Raubal, H.J. Miller, and S. Bridwell. User-centred time geography for location-based services. *Geografiska Annaler Series B Human Geography*, 86(4):245–265, 2004.
12. T. Reichenbacher, P. Crease, and S. De Sabbata. The concept of geographic relevance. In *Proceedings of the 6th International Symposium on LBS & TeleCartography* , 2009.
13. T. Reichenbacher and S. De Sabbata. Geographic relevance: different notions of geographies and relevancies. *SIGSPATIAL Special*, 3(2):67–70, 2011.
14. W.R. Tobler. A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46:234–240, 1970.
15. J. Urbano, J. Morato, M. Marrero, and D. Martín. Crowdsourcing preference judgments for evaluation of music similarity tasks. In *ACM SIGIR Workshop on Crowdsourcing for Search Evaluation*, pages 9–16, 2010.
16. M. Van Kreveld, I. Reinbacher, A. Arampatzis, and R. Van Zwol. Multi-Dimensional Scattered Ranking Methods for Geographic Information Retrieval. *Geoinformatica*, 9(1):61–84, 2005.
17. Z. Zhang, A.L. Gentile, and F. Ciravegna. Recent advances in methods of lexical semantic relatedness–a survey. *Natural Language Engineering*, 2012.

Exploring the Addressing Contexts

Ali Javidaneh[1*] and Farid Karimipour[1]

[1] Department of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Tehran, Iran
{ajavidaneh,fkarimipr}@ut.ac.ir

Abstract. The addressing (also called geocoding) as an applied spatial service should equipped with context-awareness capabilities in order to support context-based ubiquity. This paper is going to show how people find the addresses in various situations (different contextual values). To do so, having presented a context definition, addressing contextual parameters are identified using activity theory and then their values are set by filling and analyzing some questionnaires using Tehran as a case study.

Keywords: address matching, context awareness, activity theory

1 Introduction

Addressing (also called geocoding) is a form of applied spatial analysis that is frequently used in everyday life. Sometimes there is no addressing standard whatsoever, and addresses are expressed in natural languages. An interesting example is Iran, where people express addresses as a sequence of spatial elements (e.g. streets, squares, landmarks, etc.), starting from a known element. Such textual addresses not only specify the destination, but also tell how to reach the location. In this case, addressing may benefit from the advantages of context awareness, as well as other services and other forms of spatial analysis. By identifying the context parameters of addressing, human behavior for the process of finding addresses can be simulated in the form of a context-aware addressing agent. Such an agent can also evaluate different environments from an addressing point of view; i.e. after the design and implementation of such an agent, it can be used in different environments to determine the addressing shortages of that environment.

This paper aims to explore addressing contexts in order to establish how people find addresses in various situations. Such knowledge could then be used for the design and implementation of an addressing agent. Exploring these addressing contexts means identifying the associated contextual parameters, and then setting their values.

Before identifying the various contexts in every domain, however, there is an important question that must first be answered: “what is a context, and what are its types?”

Published in the Proceedings of the 1st International Workshop on Context-Awareness in Geographic Information Services, in conjunction with GIScience 2014, edited by Haosheng Huang, Jürgen Hahn, Christophe Claramunt, Tumasch Reichenbacher, CAGIS 2014, 23 September 2014, Vienna, Austria, 2014.

To answer that, Section 2 of the paper will discuss the definition of "context", and its categories. One claim that is often made is that something is context when it is used to adjust the interactions between people and activities. "Activity" can therefore be considered as the main core of context specification, which is why we propose the use of activity theory for the identification of contextual parameters. Section 3 of the paper will present a brief introduction to activity theory and its use in context identification.

In Section 4, the contextual parameters of addressing are identified using activity theory, marking the first step towards exploring addressing contexts. The values of those parameters are then set in the step by analyzing data from questionnaires, using Tehran as a case study.

In summary, the main aims of this paper are:

1. Identifying the contextual parameters of addressing via activity theory
2. Setting the values of identified contextual parameters via data retrieved from questionnaires in Tehran as a case study

2 Context definition and categorization

Context awareness may increase the capabilities and the efficiency of a system or service. A context-aware system can serve its clients by delivering more relevant services, and by decreasing the users' interference. The main goal of a ubiquitous or context-aware system is the support of context-based ubiquity, i.e. serving users according to their contexts.

Schilit, *et al.* first defined context as "interrelated conditions in which an object exists or occurs" [1]. This definition shows that context is a mutual relation between different conditions in which an agent exists or occurs.

Subsequent proposed definitions of context usually came in the form of examples [2-6] or synonyms [7-9], which were commonly project-oriented or application-based. In other words, these definitions suffered from two main shortcomings: a lack of generality, and incompleteness [10]. Moreover, these definitions were mostly user-oriented and couldn't represent environmental conditions. This led Lieberman, *et al* to propose a system-oriented definition of context as "context can be considered to be everything that affects the computation except the explicit input and output" [11]. In their definition they replaced the user with the application and system.

Abowd, *et al.* attempted to define context in a way that solves the mentioned problems of context definition. They defined context as "any information used to determine the situation of an entity" [12]. The term "entity" in this case includes both user and system, thereby defining context as a relation of user and system instead of a user-based or system-based definition. The definition is universal enough, but the terms "any information" and "determining the situation" make it imprecise. Zimmermann, *et al.* therefore tried to compensate for the imprecision of Abowd, *et al.* definition by enumerating five categories of contexts: "individuality", "time", "space", "activity" and "relations" [10]. Their attempt at defining context categories shows the importance of context categorization for the design and implementation of context-aware systems, and so many efforts have been made since to categorize contexts. Poslad defined three main categories of context [13]:

1. Physical environment context: contexts related to the physical world, such as space, time, temperature, etc.
2. User context: interactions in systems are roughly dependent on user situation. A user's personality, favorites, needs, etc. tune the system's services.
3. Virtual environment context: each component of the distributed system is aware of existing services of the system.

Schilit, *et al.* categorized context through three main questions - "where are you?", "with whom are you?", and "what resources are near you?" [14] - while Ryan, *et al.* categorized context into four categories of "space", "environment", "identity" and "time" [8]. The former categorization does not comment further on "time" and "activity".

Abowd, *et al.* claimed that environment is equal to context, and then replaced "environment" with "activity" in Ryan, *et al.* categorization [12]. Schmidt also proposed six main categories of context: "user", "social environment", "task", "condition", "infrastructure", and "space" [15]. While these two categorizations are conceptually close to each other, they have some trivial differences. One of them is that "time" in the Schmidt categorization has not been expressed explicitly; another is that the wide concept of "environment" in the Abowd, *et al.* categorization is replaced with "space".

One important definition of context was proposed by Dourish [16]. He considered two main aspects of context: the representational aspect, and the interactive one. Then he claimed that the correct aspect is the one that notices the interaction of objects and not just the representation of context. Winograd also claimed that something is context when it is used to adjust the interactions between people and activities [17]. Activity seems to be playing an important role in the field of context identification. We therefore propose the use of activity theory in order to determine contextual parameters.

3 Identifying contextual parameters using activity theory

According to the role of activity in context identification, it seems that activity theory can be an efficient tool in identifying contextual parameters. In this section we will give a brief introduction to activity theory, and then present its usefulness in context identification.

3.1 Activity theory

Activity theory is a suitable descriptive tool for the analysis and understanding of activity independent of a specified domain of application [18]. In this theory, activity has a hierarchical structure. Each activity is divided into several goal-oriented actions. The core of an action is formed by "subject", "object", "society", and "outcome". A “subject” interacts with an “object” in a “society”. These interactions are formed via “division of labor”, “rules”, and “tools”. To identify the components of an activity, first it must be divided into actions, and then the action components are identified according to figure 1.

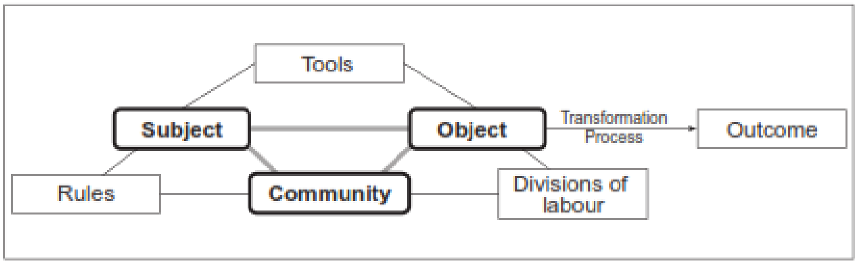


Fig. 1. The framework of activity theory

Context-aware systems designed to simplify human activities must reflect the key components of these activities. This theory can therefore be used to identify contextual parameters in any domain. Some attempts have already been made to use this theory to identify the contextual parameters of some domains [19-22], however none of them model the environmental contexts, such as space, light, and sound.

3.2 Action as the analysis unit

In activity theory, activities are long-term structures composed of short-term ones, known as goal-oriented actions [23]. If activities are considered as user tasks, then actions will be the steps within, and the respective contexts are different [24].

For the activity levels, the ones which are to be supported by the context-aware system have to be identified. In doing so, the world schema can be restricted and the tasks can be made manageable. This is helpful in identifying the activities’ contexts. Actions have their own goals, and these sub-goals in turn shape the main goal of the activity. Tasks performed under the same action have the same contextual parameters, so here we use action as a unit to identify the context.

3.3 Contextual parameters of each action

The specifications of a physical environment – such as space, weather, humidity, etc. – affect human activities therein. That environment however is not modeled by activity theory. The theory therefore needs to be extended in order to add the physical environment wherein the activities occur into the theory framework, as shown in figure 2 [25].

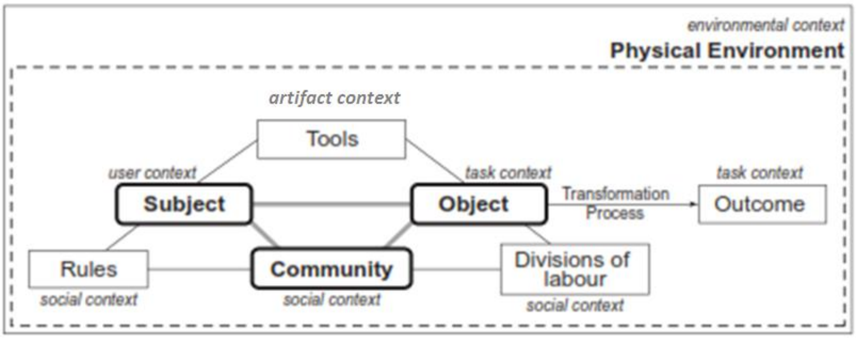


Fig. 2. The extended framework of activity theory [25]

Context categories and activity theory components can be mapped to each other as shown in table 1 [25].

Table 1. Map of activity theory components to context categories

Components of Activity Theory	Context Categories
Subject	User context
Object, Outcome	Task context
Community, Rules, Divisions of Labor	Social context
Tools	Artifact context
Physical Environment	Environment context

Our world has a rich information content, and no system can model the contexts on the fly [19]. This means that when we design a context-aware system, we should ensure that the contextual parameters relate to the domain of activity; or, in other words, that they relate to the action goals. The extended activity theory helps us to do so.

4 Identifying the addressing contextual parameters and setting their values

In this section we want to (1) identify the contextual parameters of addressing via activity theory, and (2) set the values of identified contextual parameters via data derived from some questionnaires in Tehran as a case study.

4.1 Identifying the contextual parameters of addressing through activity theory

The four main steps of addressing are defined as parsing, standardization, correction, and matching [26]. Since an addressing agent acts individually, the three components of activity theory - "society", "division of labors", and "rules" – can be disregarded. The components of addressing activity and its related contextual parameters can therefore be mapped to each other as follows:

Table 2. Identifying the contextual parameters of addressing through activity theory

	Parsing and standardization	Correction and matching
Subject	**Addressing agent** Relevant context: user information such as age, gender, etc.	**Addressing agent** Relevant context: user information such as age, gender, etc.
Object, outcome	**Determination of address components** Relevant context: none	**Positioning the address** Relevant context: none
Division of labors, society and rules	Since addressing agents act individually, the three components of "society", "division of labors", and "rules" can be disregarded	
Tools	**The defined grammar of addresses** Relevant context: constant geo-names and spatial relations	**Spatial search in map using spatial and non-spatial relations of address** Relevant address: none
Physical environment	**The route defined by the address** Relevant context: the physical characteristics of the route defined by the address such as slope, direction, landmarks, etc.	**Map** Relevant context: none

4.2 Setting the values of identified contextual parameters (case study: Tehran)

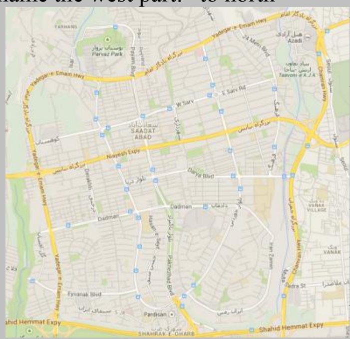
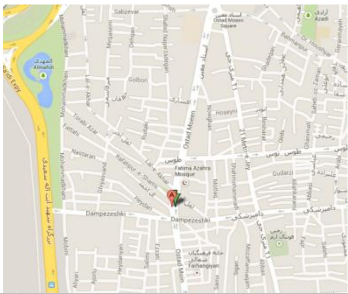
The values of contextual parameters for addressing that we identified in the previous subsection can be set by asking some people about their behavior and practice. For that purpose, we designed a questionnaire with two main parts. The first part asks about the "subject" specifications as shown in table 3.

Table 3. "Subject" specifications

Gender	*Age*	*Education level*
□ male □ female	………. years	□ Diploma □ Bachelor of Science □ Master of Science or higher

The second part asks about the subject's behavior and practices, based on some hypotheses about the "physical environment" of the addressing process. The main components of that physical environment that we asked about were "direction", "topology", "slope", "distance", and "landmarks". In reality, however, we already had some suppositions about the roles of these components in addressing, and we tested theses suppositions by asking people about their behavior and practices in addressing via the questionnaire. The hypotheses and the related questions/answers as well as the results are shown in table 4. The questionnaire was answered by about 20 people with different specifications (table 3).

Table 4. Summary of hypotheses, questions/answers, and results

Context	Hypothesis	Q/A	Result
Direction	When the arc radius becomes longer, the direction is lost. In other words, network direction is preferred to geographical direction.	On "Yadegar" highway, I prefer to name the west part: "to north" 	The subjects that drive on this highway without seeing the map usually didn't understand this directional change, i.e. the hypothesis has been confirmed. After seeing the map, however, they believe that the directional change must be noted.
Topology	Environmental information is usually formed in people's minds, topologically.	I didn't mention the break in "Ostad Moein" avenue. 	The answers of almost all of the subjects confirm the hypothesis.
Slope	In addressing, slope can be preferred to direction.	To indicate a passage to the right with a negative slope, I prefer to say "go up" instead of "go right"	The answers of almost all of the subjects reject the hypothesis.

Distance	The distance between two landmarks seems smaller than their actual distance.	Compare two distances: 1. University of Sharif to Azadi Square (a path with a landmark) 2. Enqelab Square to Valiasr intersection (a path without a landmark) Note: the two distances are same.	The answers of almost all of the subjects confirm the hypothesis.
Landmark	In addressing, referring to landmarks is preferred to referring to the names of passages.	In addressing I prefer to use landmarks instead of the names of passages.	The answers of almost all of the subjects reject the hypothesis.

The results set the values of the identified contextual parameters for addressing in Tehran as follows:

1. Network direction is preferred to geographical direction. It should be noted that some people change their answers after seeing the map. This shows that map use affects spatial cognition.
2. Spatial cognition is more topological than metrical. This result was expected and was confirmed by the answers.
3. Slope doesn't play an important role in addressing and spatial cognition. This may be because direction is recognized more explicitly than slope.
4. The distance between two landmarks seems smaller than their actual distance.
5. In cities such as Tehran, which have some clearly defined geo-names for city components, people prefer to use the names of passages instead of landmarks.

5 Conclusions

Addressing is an important spatial service in which context plays a key role. To simulate the addressing behavior of people or to design a context-aware addressing agent, one needs to identify the contextual parameters of addressing. The main contributions of this paper are (1) identifying the contextual parameters of addressing through activity theory and (2) setting the values of identified contextual parameters by analyzing questionnaires in Tehran as a case study. The values of identified contextual parameters may show how people spatially cognize the environment and use their spatial knowledge for addressing. Knowing the answers to these questions may lead to the design of some cognizing agent for addressing that can simulate human spatial behavior.

References

1. Schilit, B.N., Theimer, M.M.: Disseminating active map information to mobile hosts. Network, IEEE 8, 22-32 (1994)
2. Brown, P.J.: The stick-e document: a framework for creating context-aware applications. Electronic publishing 8, 259-272 (1995)
3. Franklin, D., Flaschbart, J.: All gadget and no representation makes jack a dull environment. In: Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments, pp. 155-160. (1998)
4. Ward, A., Jones, A., Hopper, A.: A new location technique for the active office. Personal Communications, IEEE 4, 42-47 (1997)
5. Rodden, T., Cheverst, K., Davies, K., Dix, A.: Exploiting context in HCI design for mobile systems. In: Workshop on human computer interaction with mobile devices, pp. 21-22. Citeseer, (1998)
6. Hull, R., Neaves, P., Bedford-Roberts, J.: Towards situated computing. In: Wearable Computers, 1997. Digest of Papers., First International Symposium on, pp. 146-153. IEEE, (1997)
7. Brown, P.J., Bovey, J.D., Chen, X.: Context-aware applications: from the laboratory to the marketplace. Personal Communications, IEEE 4, 58-64 (1997)
8. Ryan, N.S., Pascoe, J., Morse, D.R.: Enhanced reality fieldwork: the context-aware archaeological assistant. In: Computer applications in archaeology. Tempus Reparatum, (1998)
9. Dey, A.K.: Context-aware computing: The CyberDesk project. In: Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments, pp. 51-54. (1998)
10. Zimmermann, A., Lorenz, A., Oppermann, R.: An operational definition of context. Modeling and using context, pp. 558-571. Springer (2007)
11. Lieberman, H., Selker, T.: Out of context: Computer systems that adapt to, and learn from, context. IBM Systems Journal 39, 617-632 (2000)
12. Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., Steggles, P.: Towards a better understanding of context and context-awareness. In: Handheld and ubiquitous computing, pp. 304-307. Springer, (2001)
13. Poslad, S.: Ubiquitous computing: smart devices, environments and interactions. John Wiley & Sons (2011)
14. Schilit, B., Adams, N., Want, R.: Context-aware computing applications. In: Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on, pp. 85-90. IEEE, (1994)
15. Schmidt, A.: Ubiquitous computing-computing in context. Lancaster University (2003)
16. Dourish, P.: What we talk about when we talk about context. Pers Ubiquit Comput 8, 19-30 (2004)
17. Winograd, T.: Architectures for context. Human-Computer Interaction 16, 401-419 (2001)

18. Barthelmess, P., Anderson, K.M.: A view of software development environments based on activity theory. Computer Supported Cooperative Work (CSCW) 11, 13-37 (2002)
19. Kofod-Petersen, A., Cassens, J.: Using activity theory to model context awareness. Modeling and Retrieval of Context, pp. 1-17. Springer (2006)
20. Kaenampornpan, M., O'neill, E.: Modelling context: an activity theory approach. Ambient Intelligence, pp. 367-374. Springer (2004)
21. Cassens, J., Kofod-Petersen, A.: Using Activity Theory to Model Context Awareness: A Qualitative Case Study. In: FLAIRS Conference, pp. 619-624. (2006)
22. Kaenampornpan, M., O'Neill, E., Ay, B.: An integrated context model: Bringing activity to context. In: Proc. Workshop on Advanced Context Modelling, Reasoning and Management. (2004)
23. Nardi, B.A.: Context and consciousness: activity theory and human-computer interaction. Mit Press (1996)
24. Chen, Y., Atwood, M.E.: Context-centered design: bridging the gap between understanding and designing. Human-Computer Interaction. Interaction Design and Usability, pp. 40-48. Springer (2007)
25. Huang, H., Gartner, G.: Using activity theory to identify relevant context parameters. Location Based Services and TeleCartography II, pp. 35-45. Springer (2009)
26. Yang, D.H., Bilaver, L.M., Hayes, O., Goerge, R.: Improving geocoding practices: evaluation of geocoding tools. Journal of medical systems 28, 361-370 (2004)

Data Driven Contextual Knowledge from and for Maritime Situational Awareness

Alfredo Alessandrini, Pietro Argentieri, Marlene Alvarez Alvarez, Thomas Barbas, Conor Delaney, Virginia Fernandez Arguedas, Vincenzo Gammieri, Harm Greidanus, Fabio Mazzarella, Michele Vespe, Lukasz Ziemba

European Commission, Joint Research Centre (JRC), Institute for the Protection and Security of the Citizen (IPSC), Maritime Affairs Unit, Via Enrico Fermi 2749, 21027 Ispra VA, Italy

Abstract: Maritime Situational Awareness (MSA) is nowadays of paramount importance for safety and security. Many newly developed surveillance technologies have reached operational maturity, leading to a level of information that often enables the detection and tracking of vessels of interest at global scale. Nevertheless, the knowledge of the spatial distribution of vessels at sea is not sufficient to achieve useful understanding of their activities, to project their position into the future or to detect low-likelihood behaviours. In this paper we analyse knowledge discovery aimed at extracting contextual information of human related activities from vessel positioning data, and subsequently applying that information for increased understanding. These data driven approaches enable a high level of situational awareness that is often difficult to access, hard to update or otherwise challenging to unveil.

1 Introduction

The aim of surveillance and situational awareness research in the maritime domain is to give operational authorities a better picture of what is happening at sea. This is required by applications in the fields of maritime safety and security, border control, marine environment and pollution monitoring, fisheries control, law enforcement, defence and customs. As defined in [1], “maritime situational awareness is the effective understanding of activity associated with the maritime domain that could impact the security, safety, economy, or environment […]”. More precisely, situational awareness “involves far more than merely being aware of numerous pieces of data. It also requires a much more advanced level of situation understanding and a projection of future system states in light of the operator’s pertinent goals” [2]. The knowledge of spatial distribution of vessels can thus be sufficient for maritime surveillance, but necessarily needs to be enriched with additional information to achieve situational awareness. The same vessel manoeuvre can appear either normal or anomalous depending on the specific context in which it is observed. In this paper, the *context* is defined as any piece of information that can be used to better understand the scene and characterise relevant vessel behaviours or intents, ultimately aiding Maritime Situational Awareness applications. An example of contextual information is the set

Published in the Proceedings of the 1st International Workshop on
Context-Awareness in Geographic Information Services, in conjunction with
GIScience 2014, edited by Haosheng Huang, Jürgen Hahn, Christophe Claramunt,
Tumasch Reichenbacher, CAGIS 2014, 23 September 2014, Vienna, Austria, 2014.

of patterns, activities, traffic *etc.* representing the behavioural "normality" in the area where a vessel is moving. The extent to which a vessel track or behaviour aligns or deviates from such contextual information can be thought of as a measure of the degree of behavioural conformity or anomaly respectively. Moreover, the same contextual information can be used to infer future vessel states. Scene understanding through contextual information and projection of the current state into the future are therefore key elements of situational awareness.

Today, the main challenge of maritime situational awareness is to aggregate large amounts of heterogeneous data and transform them into useful and reliable information to be used by operators and relevant stakeholders in the decision making process. This challenge is made more difficult by the need to achieve a global maritime situational awareness due to the emerging requirement of detecting unlawful activities that can have consequences even if taking place in international waters.

In tracking systems applications research efforts are concentrating on the integration of context information to improve overall performance (see *e.g.* context aware video tracking [3] and knowledge based radar detection tracking and classification [4]). More specifically, in the field of maritime surveillance and situational awareness, there is also an emerging interest for context awareness, necessary to achieve low-likelihood behaviour detection, destination prediction and reconstruction of vessel tracks [5- 9].

The paper aims to provide an introduction to maritime situational awareness by giving practical details on data integration and fusion, along with a description of current implementations and future developments to improve the characterisation of activities at sea by transforming low level positioning information into contextual features.

2 Maritime Situational Awareness Data

Maritime situational awareness data can relate to vessels and their attributes (position, identification, history, etc.) and to contextual geographically-linked information.

Vessel positioning data and static information

Information on the spatial distribution of vessels at sea can be grouped into *self-reporting* or *observation-based* depending on the way such data are acquired. In particular (see **Fig. 1**):

- *Self-reporting positioning data*, also often referred to as "cooperative", are transmitted by the vessel in the vessel proximity (Automatic Identification System – AIS – for collision avoidance) or to competent authorities (Long Range Identification and Tracking – LRIT – for security and safety, Vessel Monitoring System – VMS – for fisheries monitoring). The technical specifications of such systems are regulated at national or international level and the relevant tracking capabilities are characterised by refresh rate, quantity and types of vessels under regulation (vessel coverage) and, sometimes, spatial coverage and transmission delay. The information content of self-reporting positioning data, besides the state vector and other kinematic information, may also include voyage related and static information about the ship.

- *Observation-based positioning data* are instead collected by active or passive sensors providing detection capabilities that vary depending on specific parameters (*e.g.* resolution, spatial coverage, update rate, latency), conditions (*e.g.* sea state/metoc data) and target properties (*e.g.* size and orientation). Observation-based data include space-based (Earth Observation - EO) Synthetic Aperture Radar (SAR), EO - Optical Images, coastal/mobile ground based radars or mounted on Maritime Patrol Aircrafts (MPA) or Remotely Piloted Aircrafts (RPAs). Additional sources of information for wide area surveillance that are topics of promising research are also HF radars, and airborne systems, passive radar applications and radiolocation of emissions.

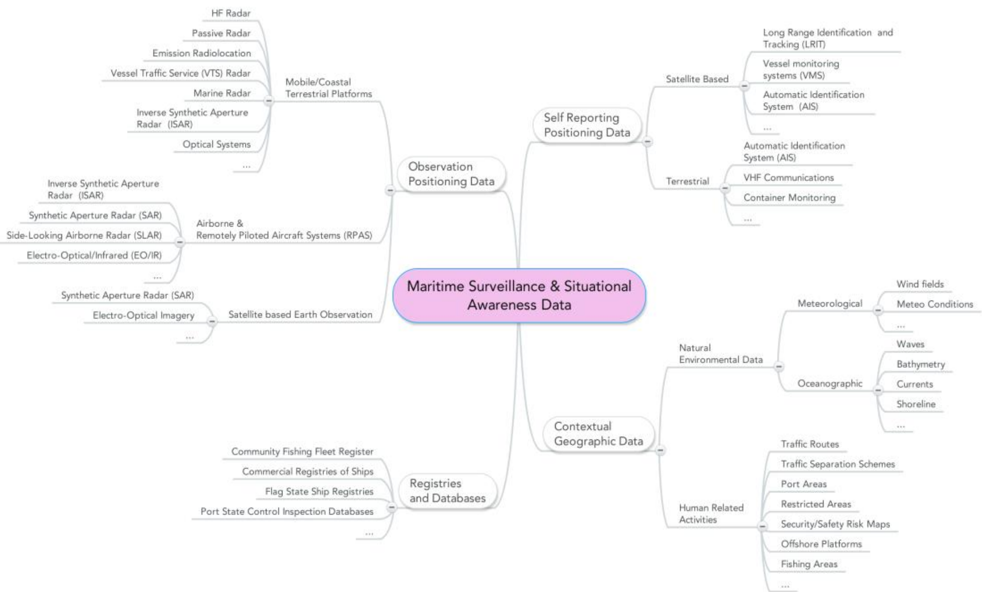


Fig. 1. Maritime surveillance positioning (self-reporting and observation-based), registries/databases and contextual geographic data for Maritime Situational Awareness.

- *Information Registries and Databases***:** vessel registries and databases contain information linking a ship identity to details about its structure, construction, appearance, history, management, and safety/security inspections. Such information is static, meaning that it can be thought of as time unvarying or slowly changing with respect to the vessel current position and motion. Vessel registries and database therefore provide complementary data to positioning information. After the ship inspection of Port State Control (PSC), according to PSC regional agreements (Paris MoU, Tokyo MoU, Black Sea MoU, etc.), if deficiencies are found, actions commensurate to the nature of the detected issues are taken. For instance, PSC may decide to inform the relevant Flag State, request the Master to rectify the deficiency before the departure or at next port of call. This information is contained in vessel registries, which can be therefore used for increasing the maritime situational awareness.

Contextual Geographic Data
The context in which vessels move and operate has to be taken into account in order to accurately understand and monitor activities at sea. Two main categories of contextual geographic information have specific relevance to security and safety aspects of the maritime domain:

- *Natural Environmental Data:* The global marine sector (marine sciences and maritime) generates and consumes enormous volumes of data every day. Some of this data such as meteorological (wind fields and generic weather conditions) and oceanographic data (currents, bathymetry, shoreline etc.) are essential for safety of navigation and risk assessment applications. Recent developments (*e.g.* see [10,11]) have improved access to worldwide forecasts of ocean conditions and observation data.

- *Human Related Activities:* traffic management or routing systems, off-shore installations, port areas and facilities, risk or restricted areas etc. represent crucial information for both maritime safety and security. This contextual information is often difficult to source or access and is sometimes in the form of text (*e.g.* international regulations) that is often hard to keep updated. In other cases like for unregulated routes, this information is unavailable. However, by analysing common behaviours it is possible to extract many elements of human related activities. This is further described below.

3 Data Fusion

As can be seen from **Table 1**, there is no ideal sensor or technology that does not present performance limitations in any of the indicators and that can be used as a sole source of information for all surveillance applications in the maritime domain. Moreover, there are specific applications that cannot be properly resolved even by using all realistically available data. This is the case, for instance, for irregular migration, where small boats need to be detected and persistently tracked in open sea, calling for innovative sensors and technologies. Nevertheless, it is clear that a level of integration and fusion can overcome the limitations of single self-reporting and observation-based systems.

		Spatial Coverage	**Vessels Coverage**	**Probability of Detection (P_d) & False Alarm (P_{FA})**	**Refresh rate / Tracking capabilities**	**Data latency**
Self-reporting systems	**Terrestrial - AIS**	VHF propagation, nominally line-of-sight	Only SOLAS Regulation V/19-2 vessels	All vessels covered and in range are detected. P_{FA} limited by spoofing	Always adequate by design	Virtually no latency
	Satellite -AIS	Virtually global	Only SOLAS Regulation V/19-2 vessels	*Pd* can depend on ship. De-collision algorithms needed in busy areas. P_{FA} limited by spoofing	Depending on # satellites in constellation	Depending on the Latitude/visibility of ground station
	LRIT	Virtually global	Only SOLAS Regulation V/19-1 vessels	All vessels covered are detected. P_{FA} limited by spoofing	Every 6 hours, can be polled any time	Network-related (Sat communications)
	VMS	Virtually global	Fishing vessels in excess of 12 m (in EU)	All vessels covered are detected. P_{FA} limited by spoofing	Up to hourly rate, can be polled any time	Network-related (Sat communications)
Observation positioning sensors	**Coastal / Mobile Radar**	Nominally 20 NM from the coast	All in range, depends on P_d	Depends on sea state, Radar Cross Section (RCS), frequency and polarisation	Continuously scanning radar	Virtually no latency
	EO Synthetic Aperture Radar (SAR)	Virtually global	Depends on P_d	Depends on sea state, RCS, frequency and polarisation	Latitude dependent, reduced by sun-synchronous orbits	Depending on the Latitude/visibility of ground station
	EO-Optical	Virtually global	Depends on P_d	Depends on sea state, target size and cloud cover	Latitude dependent, usually reduced by sun synchrounous orbits	Depending on the Latitude/visibility of ground station
	RPA/Airborne radar systems	Limited by operational costs	All in range, depends on P_d	Depends on sea state, RCS, frequency and polarisation	Continuously scanning radar if aircraft is present	Virtually no latency

Table 1. "Traffic light" table of the main operational observation-based and self-reporting positioning systems for maritime situational awareness.

However, heterogeneous sensors and technologies intrinsically present different sampling time, data latency (*i.e.* time lag between the data collection and the time they are made available), errors and uncertainties. A data fusion strategy together with practical details of self-reporting and observation-based positioning data association to produce the maritime picture can be found in [12].

4 Contextual Knowledge Discovery

Knowledge discovery can be applied to a time series of maritime pictures to extract human related activities contextual information. Data fusion is a necessary step to improve the knowledge discovery process: a track that is enriched by additional observations or self-reporting positions leads to a more complete maritime picture and therefore to a more meaningful and accurate knowledge.

The output of knowledge discovery can then be used as a reference to complete missing information, to project the situation into the future, and to detect low-likelihood behaviours as summarised in **Fig. 2**.

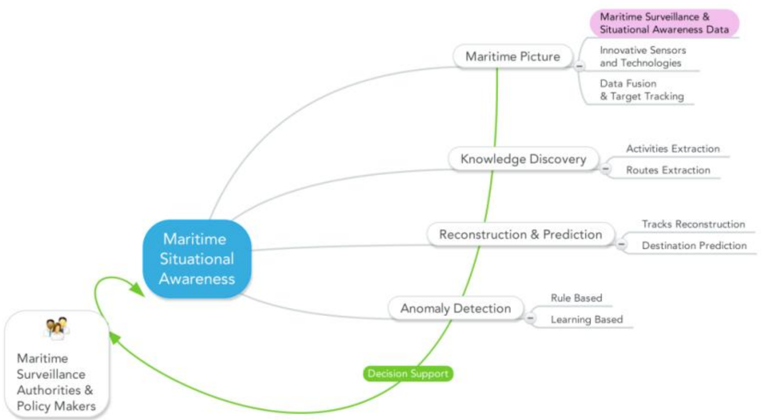


Fig. 2. Decision support tools for Maritime Situational Awareness built on the time series of maritime pictures.

4.1 Traffic Density

Traffic density maps are the first type of layers to efficiently represent valuable contextual information. As an example, main traffic routes and concentrations near ports off West Africa can be observed in **Fig. 3**. Self-reporting data have been fused to form tracks that progressively shape and highlight the highest vessel density areas. The routes are however mixed and it is difficult to isolate specific itineraries.

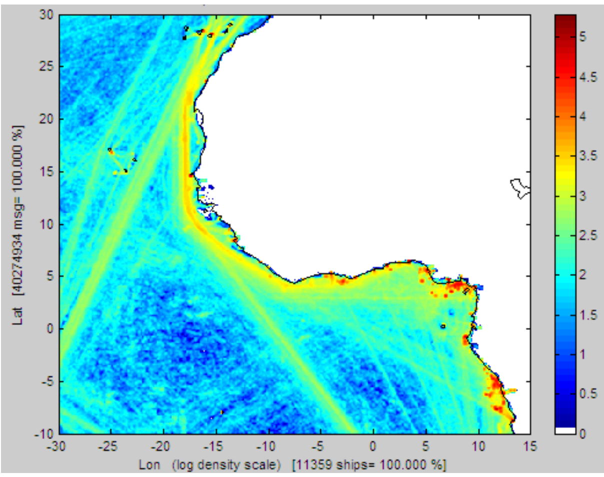


Fig. 3. Vessel traffic density using 6-month ship reporting data from Satellite-AIS (exactEarth, SpaceQuest, LuxSpace, Orbcomm, Norwegian Coastal Administration) over West Africa.

4.2 Patterns of Interest

A higher level of contextual description can be achieved by applying spatial data mining to positioning data, allowing the decomposition of the maritime traffic into sets of patterns obtained by clustering vessel trajectories as shown in **Fig. 4**.

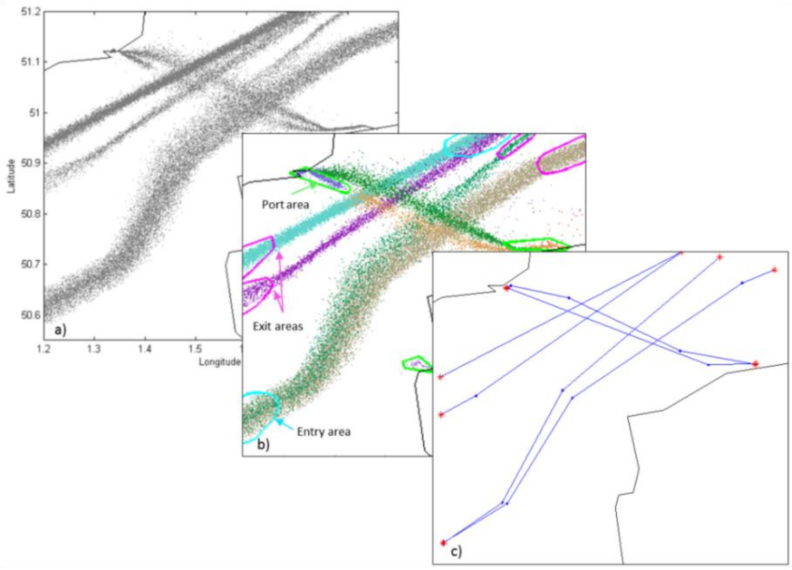


Fig. 4. Traffic route extraction from two weeks of AIS data in the Dover Strait (a), main traffic patterns clustered (b) and network based representation (c), after [13].

As further discussed in [13], first Entry/Exit areas are automatically detected as clusters of points where vessels are first/last seen. Ports and stationary areas are automatically extracted by grouping points where vessels consistently reduce their speed. The trajectories connecting Entry-to-Exit, Entry-to-Port, Port-to-Port or Port-to-Exit areas are then grouped to form the set of patterns in the scene, each of them described in terms of velocity, ship-type and travel time distributions. Such contextual information can be used as basis for situational prediction and anomaly detection.

4.3 Mapping of Fishing Activities

From the analysis of historical positioning data it is also possible to map generic activities on the basis of specific vessel motion models associated. As an example, by using positioning data it is possible not only to separate steaming (high speed and low heading variability) from stationary (nearly zero speed over ground) or fishing (low speed and consistent heading variability) behaviours for each the tracked vessels, but also to cluster the collective behaviours related to all vessels in an area of interest. The result is a map that gives indications on the fishing activities spatial and intensity distribution as shown in **Fig. 5**. As reported in [14], the concentration of these points is strictly linked to the fishing effort in the area of interest, representing a valuable contextual information layer that can be used in support to decision makers.

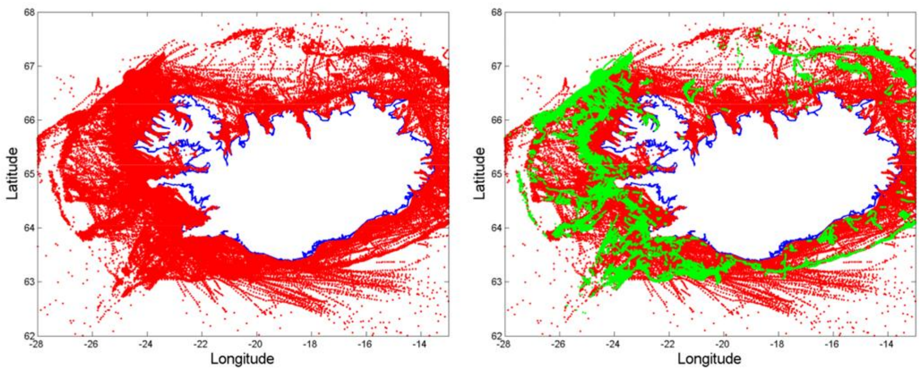


Fig. 5. One-month AIS positioning data from the MSSIS network in Icelandic waters (red) and probable fishing activities automatically extracted from the same data (green) ultimately leading to the mapping of fishing footprints, after [14].

5 Contextual Knowledge Application

Contextual knowledge and normal behaviour models can be useful to human operators as visualisation and analytical support layer [15], but can also be integrated into automatic processing chains for situation prediction, track reconstruction or anomaly detection.

5.1 Destination Prediction and Route Reconstruction

The patterns of interest introduced in Section 4.2 can be used to implement automatic *i*) destination prediction, *i.e.* the route is used to predict the future position and the eventual destination of a vessel, and *ii*) route reconstruction from a relatively sparse set of positions. As an example, in **Fig. 6** a track is reconstructed from two observations belonging to the same vessel, reflecting the most probable compatible route.

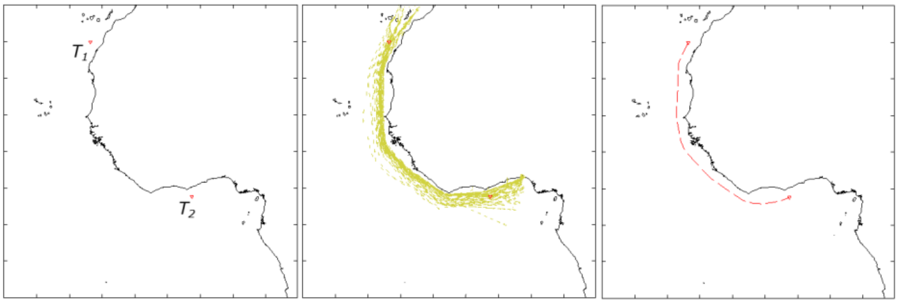


Fig. 6. Positioning data at time T_1 and T_2 with several days of missing information in between (left) in West Africa. The observations are found to belong to the underlying route (middle) out of a number of additional ones in the area, and the non-linear trajectory estimated (right).

Such contextual knowledge can also be used to connect observation-based positions (*e.g.* SAR ship detections), where the identity of the ships, in most of the cases, can only be estimated.

5.2 Anomaly Detection

The problem of finding anomalous behaviours can be approached either by defining specific alert rules or, alternatively, by detecting low-likelihood events. The former is achieved by using spatial ontologies [16] or by coding Subject Matter Expert (SME) knowledge into a set of logic expressions [17] such as a vessel that crosses a predefined area or when the speed exceeds a certain threshold etc. Such thresholds and polygons can alternatively be derived from the normality of traffic that is observed in the area of interest [9]. Spatial data mining and knowledge discovery processes can therefore lead to information to be used as a reference to measure and detect unexpected behaviours. Deviations from the "normality" can be highlighted for further analysis as in the example shown in **Fig. 7**.

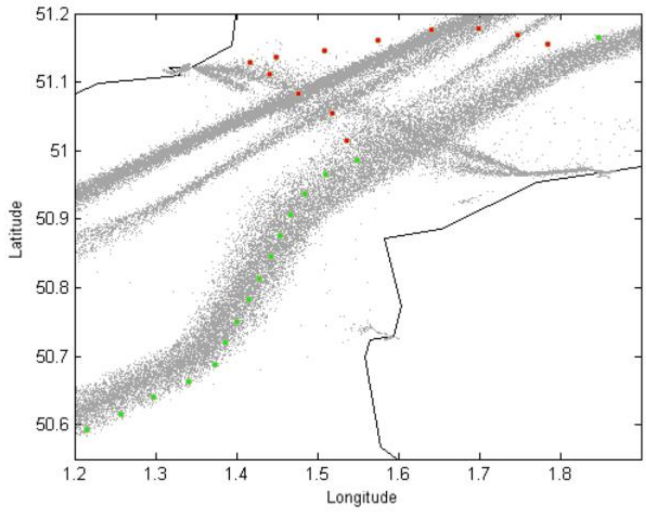


Fig. 7. A vessel track following one of the extracted eastbound patterns (green) in the Dover Strait before crossing the channel back and forth without stopping (red) and eventually realigning to the main route.

6 Conclusions

One of the goals of Maritime Situational Awareness is filtering the large amount of positioning data for operators and providing them with additional layers useful to the decision making process. Part of contextual awareness can be achieved by extracting knowledge of common vessel activities at sea from their positioning data. The advantage of such data driven approach is to discover subtle and dynamic information that can be useful to enhance the understanding of activities at sea. Ways to model, extract and use such contextual information in the maritime domain have been presented in this work, together with a set of practical examples.

Acknowledgements

The data fusion and knowledge discovery algorithms are implemented within the framework of the JRC Blue Hub project [18].

The U.S. Navy and the Volpe Centre of the U.S. Department of Transport are acknowledged for the MSSIS data.

References

1. Towards the integration of Maritime Surveillance: A common information sharing environment for the EU maritime domain, COM (2009) 538 final, 15.10.2009.
2. Endsley, M. R., "Toward a theory of situation awareness in dynamic systems" in Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1), 32-64, 1995.
3. Yang, M., Wu, Y., & Hua, G. "Context-aware visual tracking" Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(7), 1195-1209, 2009.
4. Gini, F.; Rangaswamy, M. "Knowledge based radar detection, tracking and classification", John Wiley & Sons, 2008.
5. Battistello, G.; Ulmke, M. "Exploitation of a priori information for tracking maritime intermittent data sources." Proc. 14th Int. Conf. on Information Fusion, 2011.
6. Vespe, M.; Sciotti M., Burro, F.; Battistello, G.; Sorge, S. "Maritime multi-sensor data association based on geographic and navigational knowledge.", RADAR'08, pp. 1-6, 2008.
7. George, J.; Crassidis, J.; Singh, T.; Fosbury, A. M., "Anomaly Detection using Context-Aided Target Tracking", J. Adv. Inf. Fusion, 6(1), 39-56, 2011.
8. Falchetti, S.; Alvarez, M.; T. Cokacar; Greidanus, H.; Vespe, M. "Improving Cooperative Vessel Tracking for a Maritime Integrated Surveillance Platform" In Proc. NATO RTO SCI-247 Symposium on "Port and Regional Maritime Security", Lerici, 2012.
9. Pallotta, G.; Vespe, M.; Bryan, K. "Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction." Entropy 15, no. 6: 2218-2245, 2009.
10. ERDDAP, the Environmental Research Division's Data Access Program. [Online]. Available: http://coastwatch.pfeg.noaa.gov/erddap/index.html
11. EMODNET, the European Marine Observation and Data Network, [Online]. Available: www.emodnet.eu
12. Mazzarella, M.; Alessandrini, A.; Greidanus, H.; Alvarez, M.; Argentieri, P.; Nappo, D.; Ziemba, L. "Data Fusion for Wide-Area Maritime Surveillance", In proceeding of: COST MOVE Workshop on Moving Objects at Sea, Brest, 2013.
13. Fernandez Arguedas, V.; Pallotta, G.; Vespe, M. "Automatic Generation of Geographical Networks for Maritime Traffic Surveillance", Proc. 17th Int. Conf. on Information Fusion, 2014.
14. Mazzarella, F.; Vespe, M.; Damalas, D.; Osio, G. "Discovering Vessel Activities at Sea using AIS Data: Mapping of Fishing Footprints", Proc. 17th Int. Conf. on Information Fusion, 2014.
15. Riveiro, M. J. "Visual analytics for maritime anomaly detection.", 2011.
16. Vandecasteele, A.; Devillers, R.; Napoli, A.: "A semi-supervised learning framework based on spatio-temporal semantic events for maritime anomaly detection and behavior analysis", In Proceedings CoastGIS 2013 Conference: Monitoring and Adapting to Change on the Coast, 2013.
17. Roy, J. "Anomaly detection in the maritime domain." In SPIE Defense and Security Symposium, pp. 69450W-69450W. International Society for Optics and Photonics, 2008.
18. Blue hub platform. [Online]. Available: https://bluehub.jrc.ec.europa.eu

Unsupervised Maritime Pattern Analysis to Enhance Contextual Awareness

Virginia Fernandez Arguedas[1], Giuliana Pallotta[2], and Michele Vespe[1]

[1] European Commission - Joint Research Centre (JRC)
Via E. Fermi 2749, 21020 - Ispra, Italy
[2] NATO Centre for Maritime Research and Experimentation (CMRE)
Viale San Bartolomeo 400, 19126, La Spezia - Italy

Abstract. Maritime Situational Awareness aims at monitoring maritime activities and ensuring safety and security, based on contextual knowledge. Maritime contextual information is difficult to access, resource-consuming to update and sometimes unavailable. Thus, data-driven approaches to derive contextual information are required for supporting maritime situational awareness systems. In this paper, a data-driven algorithm is proposed to extrapolate maritime traffic contextual information from real-time self-reporting data. The knowledge discovery process focuses on the detection and definition of the maritime corridors, based on the construction of maritime traffic networks. The maritime traffic network provides maritime contextual knowledge to automatically update the Maritime Situational Picture, contributing towards Maritime Situational Awareness and risk management systems evolution.

Keywords: Maritime Situational Awareness, Maritime Surveillance, Knowledge Discovery, Traffic Networks, and Anomaly Detection

1 Introduction

Maritime surveillance has become an increasingly active research field due to the broad range of activities covered and for its direct implications with security, safety, environmental and socio-economical factors. Maritime surveillance targets improving users' safety and security as well as protecting maritime traffic. In the European Union, 74% of goods imported and exported and 37% of exchanges within the Union transit through seaports [7]. Moreover, 385 million passengers pass by ports every year, converting Europe's ports on the gateways to the European continent [8]. Moreover, the total gross weight of goods handled in EU ports was estimated at 3.7 billion tonnes in 2012 [1]. Such statistics enlighten the relevance of monitoring and protecting maritime traffic.

Maritime surveillance represents a challenging research area, covering a wide range of activities, *e.g.*, irregular migration, piracy, fisheries control, traffic monitoring, etc. However, the maritime surveillance ultimate objective is to understand activities at sea, provide a maritime situational picture and project it into the future for risk prediction and assessment. Maritime Situational Awareness

Published in the Proceedings of the 1st International Workshop on
Context-Awareness in Geographic Information Services, in conjunction with
GIScience 2014, edited by Haosheng Huang, Jürgen Hahn, Christophe Claramunt,
Tumasch Reichenbacher, CAGIS 2014, 23 September 2014, Vienna, Austria, 2014.

(MSA) consists of the ongoing maintenance of the Maritime Situational Picture (MSP), based on a four-stage process including heterogeneous data acquisition and fusion, data analysis and results presentation [15]. Numerous efforts have been invested in the literature towards the maritime situational awareness, from the collection and fusion of heterogeneous data to feed the MPS [19], to results presentation [18] and the knowledge discovery process [16]. In [18], the authors proposed a methodology to filter out vessels by building normal behaviour models that human operators could use to continuously validate vessel behaviours, flagging only the patterns not matching the model. The authors based their approach on the idea that detection of unusual vessel activities is a key objective to enable maritime situational awareness in the homeland security domain. Other existing work on maritime anomaly detection include [17,**?**]. In [16], the authors contributed to maritime situational awareness by presenting an algorithm to predict future vessel behaviour on-the-fly given the current vessel location and velocity. The neurobiologically inspired algorithm uses real-time tracking information to learn motion patterns, which are used to adapt the models enabling the system to face evolving situations. In [19], a system to combine sensor-based information with context information and multi-source intelligence is presented. The fusion and analysis of the data to distinguish suspicious from normal behaviour is based on domain ontologies. Despite various efforts have been undertaken to support maritime situational awareness, this research area is still greatly unexplored.

In this paper, we propose to contribute to the Maritime Situational Picture by deriving contextual information from maritime traffic historical data. In this work we follow Dey's definition of context [9]: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves". In particular, the entities of interest are the vessels, whose intents and behaviours are analysed in a specific context, namely a routing systems or a set of maritime common patterns. Our contribution targets the detection, geographic location and characterisation of maritime corridors, *i.e.* routes with high maritime traffic density. The maritime corridors are represented by maritime traffic networks which are automatically generated based on typical vessels trajectories extrapolated from historical data. The projection of the maritime traffic networks on the geographical MSP enhances Maritime Situational Awareness (MSA) over the area of interest, that is the understanding of the activities of vessels at sea and their projection into the future. The results of the method are presented here for a particular scenario, the transport route between Gibraltar passage and the Dover Strait, with plans for extension to reach a global scale.

The remainder of the paper is organized as follows. The Maritime Traffic Knowledge Discovery process is detailed in Section 2, with its main blocks technically described in Sections 2.1, 2.2 and 2.3. Section 3 details the system evaluation. The paper ends with a discussion on the conclusions and future work in Section 4.

2 Maritime Traffic Knowledge Discovery

Maritime Situational Awareness(MSA) focuses on understanding the Maritime Situational Picture (MSP) in an attempt to monitor maritime activities and ensure safety and security. Maritime contextual information is typically difficult to obtain. In some instances, contextual information is contained in a legal document, thus its consideration is not automatic; whilst in others, contextual information imply “de facto” standards. Both situations imply a large amount of supervised work to obtain maritime contextual information and insert it in any maritime surveillance system. In this paper, we propose a data-driven knowledge discovery process to automatically derive the contextual information from vessel self-reporting positioning data, eliminating the manual interaction and allowing automatic updates of the MSP. The methodology is tested using Automatic Identification System (AIS)[3] data, nevertheless it can be extended to other surveillance sensors.

The envisaged contextual information consists in the detection, definition and location of the maritime corridors, areas with high dense traffic or high vessel visiting probability. The maritime corridors entail high interest regarding safety and security. Projecting such information in the MSP contributes towards the risk definition within the area and provides the MSP with an evolving nature. In this paper, the knowledge discovery process focuses on the detection and definition of the maritime corridors, based on the construction of maritime traffic networks.

Maritime traffic networks are directional graphs essentially built to represent maritime traffic as a set of nodes and routes in the area of interest, as a compact and holistic representation of maritime traffic. Maritime traffic networks delineate the areas with high maritime traffic, represent the changes in the vessels behaviour and indicate the directionality of the traffic. Using the maritime traffic network as the foundation for the maritime corridors definition enhances the MSP, contributing towards the completeness of the current MSP and providing information to predict the future MSP. The predicted MSP contributes towards the risk analysis and assessment in advance.

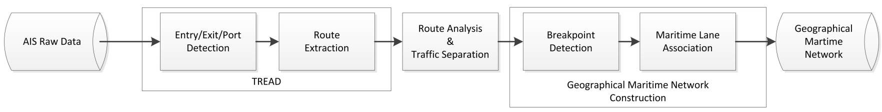


Fig. 1. Maritime Traffic Knowledge Discovery System

Maritime Traffic Knowledge Discovery proposes to perform unsupervised maritime traffic analysis for the construction of a maritime traffic network and

[3] Ships of 300 gross tons and upwards in international voyages, 500 tons and upwards for cargoes not in international waters and passenger vessels are obliged to be fitted with AIS equipment as regulated by the IMO Safety of life and sea (SOLAS) [4].

the knowledge projection into the contextual information layer. The analysis is based on processing AIS data, benefiting from the provided data regarding vessels' location, spatio-temporal evolution, etc. Our Maritime Traffic Knowledge Discovery process is envisaged as a three-stage system (refer to Figure 1). First, in Section 2.1, the waypoints and routes detection algorithm is explained. Second, the traffic separation and projection of routes in the temporal axis is further detailed in Section 2.2. Finally, Section 2.3 explains the maritime traffic network construction algorithm, presenting a fast, distributed algorithm which efficiently represents maritime traffic with a high compression rate.

2.1 Waypoint and Route Detection

Historic AIS data provides information about vessel spatio-temporal evolution and geographical location. Raw AIS data reveals the complexity of maritime traffic patterns (refer to Figure 2). In the proposed approach, maritime traffic is envisaged as a network or a directed graph composed of different elements connected and hierarchically organised. The outer layer of the network is composed of the raw maritime traffic routes and the waypoints. These patterns must be detected prior to synthetically representing maritime traffic.

In maritime traffic, three types of waypoints are considered: entry, exit and ports. Entry/Exit gates are created and dynamically update when a vessel enters/leaves the area of interest, generating "birth"/"death" events. Ports identify local ports, offshore platforms and stationary areas, detected by speed gating. Entry and exit gates depend on the area under analysis, while ports are reference points, invariant with respect to the specifically monitored area. Both Ports and Entry/Exit gates are created, expanded and merged progressively using an incremental Density-Based Spatial Clustering procedure (see [11] and [10]), where the clustering parameters are set based on specific traffic density, intensity and regularity in the area of interest. Exits, entries and ports will be used as the main nodes in the creation of the maritime traffic network (refer to Section 2.3).

Once the waypoints are derived, the positions of the vessels transiting between these waypoints are clustered and routes connecting the waypoints are formed. The route clusters include also both static information related to the type of vessels and dynamic features. The routes are activated when a minimum number of vessel trajectories is detected. Once the routes have been detected, a synthetic route is computed as an average representation of the routes. The approach reflects the knowledge discovery process of the Traffic Route Extraction and Anomaly Detection (TREAD) methodology (further described in [14], [13]).

2.2 Route Analysis and Traffic Separation

The Waypoint and Route Detection process provides information about the vessels' spatio-temporal evolution in the shape of self-contained structures defined by a set of attributes, including: timestamp, route ID, number of vessels associated to the route, MMSI number list, ship type code list, entry/exit points and route coordinates. However, the vessels' coordinates indicate a directional representation. In order to provide directional-axis-independence, the detected routes are further analysed and projected in the temporal axis. To establish the relation between the temporal axis and the route points, a temporal cell-division is proposed. The underlying idea is that temporally-consecutive coordinates would be located in a neighbouring area or cell. The cell is defined as a circular area moving over the pre-calculated synthetic routes. In each point of the synthetic route a cell is superimposed, selecting the points within its neighbourhood. The Course Over Ground (COG) related to such points are then stored sequentially. As a result, each route reveals the temporal evolution of the COG as shown in Figure 3(b).

In maritime traffic, different countries or regions establish navigation rules to prevent traffic accidents or to limit their impact on security and environment. Our scenario under analysis is the transport route between the Gibraltar passage and the Dover Strait (for further information refer to Section 3.1). This area presents several maritime Traffic Separation Schemes (TSS). TSS is a maritime regulation in high traffic density areas or confined waterways establishing non-crossing points or one-direction lanes to facilitate navigation [4]. Four TSS regulate the maritime transport in the analysed area including *Ushant*, *Finisterre*, *Cabo Roca* and *Cabo de Sao Vicent*. These specific TSS include new traffic lanes for ships carrying dangerous/pollutant cargoes in bulk (see e.g [2]).

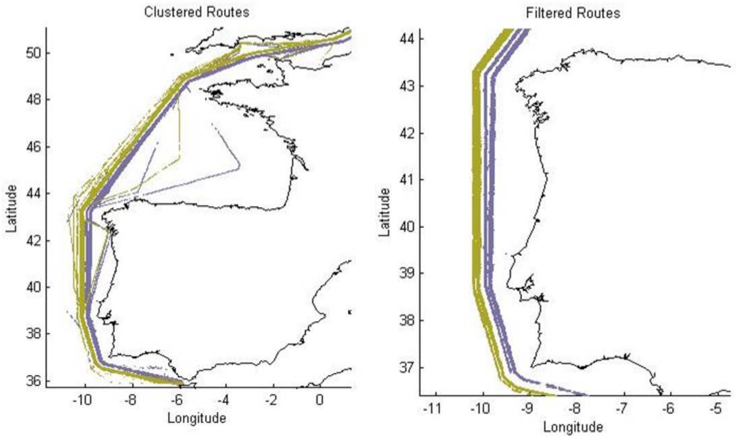
The traffic separation schemes in the analysed area imply the separation of maritime lanes with different directions, and also the creation of new lanes for a specific vessel type. Such separation reflects on the raw data collected in the area as shown in Figure 2. Thus, to consider this highly complex scenario, we propose a data-driven algorithm to independently analyse routes based on the ship type code. Each route provided in Section 2.1 is then divided into 80 subroutes [5]. The analysis of the data revealed that most of the ship type codes are associated to the inner lane (normal route) while tankers navigate in the outer lane, demonstrating the application of mentioned TSS in the area [2]. Finally, the traffic separation reflected in the data is considered in the knowledge discovery process for a more accurate maritime traffic network construction.

2.3 Geographical Maritime Traffic Network Construction

Once the routes have been unfolded in the temporal axis and separated according to the ship type code, adapting to the traffic separation schemes in the area under analysis, the maritime traffic network is constructed. The hierarchical maritime traffic representation method depicts maritime traffic as a directed graph [12]. The process is comprised of two steps repeated for each route. First, vessels' behaviour changes are detected based on detection of large and stable cumulative variations of the Course Over Ground. Such changes locate inflection

[4] Traffic separation Scheme (TSS) was approved by the International Maritime Organisation as ships routing system for collision avoidance [5].

[5] 80 different ship type codes are considered in the AIS data definition.



(a) (b)

Fig. 2. Transport Route from Gibraltar Passage to the Dover Strait. (a) Historical Maritime Traffic. (b) Extended View of the Maritime Traffic in the Portugal Coast.

points (breakpoints), which imply the division of the trajectory into segments, so called tracklets. After the breakpoints are projected in the geographical map, the Maritime Lane Manager applies a hierarchical algorithm to reconstruct the synthetic routes, based on the waypoints and breakpoints locations. In this paper, the process distinguishes between routes and subroutes, applying the method not only to the original routes but also to the ship type subroutes resulting from the traffic separation detailed in Section 2.2.

Another relevant aspect of the maritime traffic network is the percentage of each vessel type navigating through each tracklet. For instance, a vessel using the inner route close to Portugal is not a maritime anomaly, however, if that vessel is a tanker transporting pollutants, it should be signalled as an anomaly due to its disregard to the TSS existing in the area. In order to target these types of anomalies, the network has evolved towards a weighted network, where each branch reveals the a priori probability that the vessel passing by belongs to one of the vessels' types. Such probability has been statistically calculated using the AIS historical data.

Finally, once the weighted maritime traffic network is defined, the network is projected into the geographical map. The network defines the geographical areas with higher transit density, called maritime corridors. The knowledge extrapolation to contextual information services provides a new layer of information addressing risk assessment.

3 Case Study: Transport Route Between the Gibraltar Passage and the Dover Strait

3.1 Dataset & Ground Truth

The scenario under analysis is the transport route between the Dover Strait and Gibraltar Passage, revealing the transport patterns between Europe and North Africa. This area reveals one of the most complex routes defined by its high traffic density and diverse traffic separation schemes. Numerous vessels transit through this maritime corridor, many of them carrying dangerous cargoes leading to significant safety and environmental risks.

Numerous terrestrial and satellite AIS receivers are deployed along this route, providing valuable maritime traffic information for the extraction of maritime patterns. One month of terrestrial AIS data has been analysed to extract the main routes (refer to Section 2.1), and further analysed for the maritime traffic network construction and contextual information extrapolation.

3.2 Maritime Traffic Network Construction

The Maritime Traffic Knowledge Discovery process performs unsupervised maritime traffic pattern extraction and directional graph traffic representation. The proposed approach aims at (i) detecting areas with high maritime traffic density to define a risk-level contextual layer, and (ii) building directional graphs to efficiently represent maritime traffic. Section 2.3 details the maritime traffic network discovery process from historical AIS data based on the behaviour changes in the vessels' spatio-temporal trajectory.

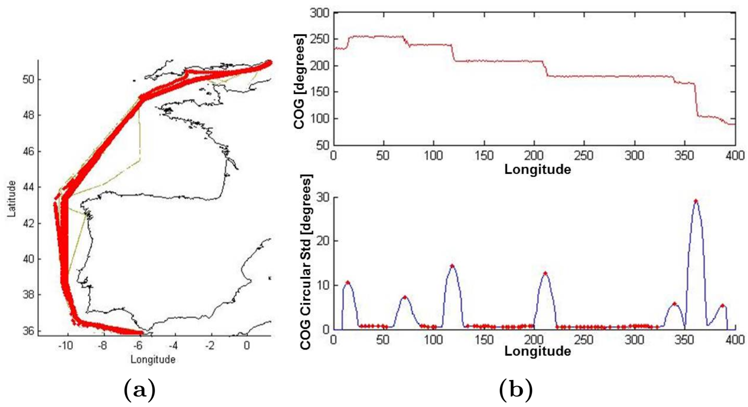
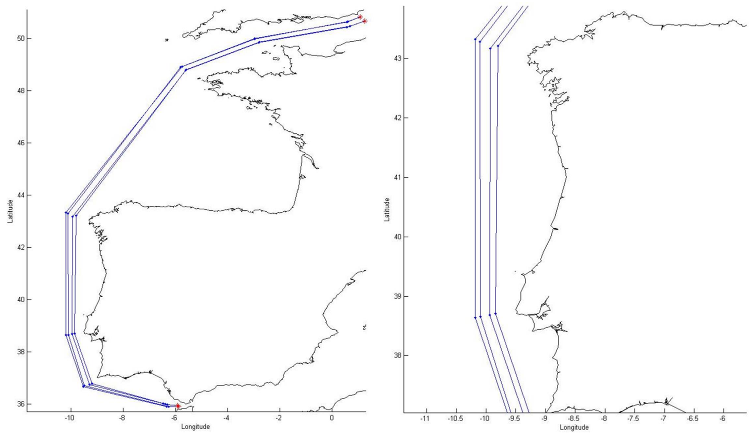


Fig. 3. Breakpoint detection on a sample route. (a) Route example showing all the raw routes. (b) COG evolution of the route and COG circular standard deviation, where the peaks relate to the vessels' behaviour changes revealed in the geographical route.

The proposed case study was selected for its high complexity regarding traffic density, transport route of pollutants, complex routes (vessel longitudinal and latitudinal directionality) and diverse traffic separation schemes. The raw data (refer to Figure 2) was used to extract routes and waypoints building the first layer of the maritime traffic network. The routes were further analysed and separated in “subroutes” to commit to the traffic separation schemes existing in the analysed area. The resulting routes passed to the *Maritime Traffic Network Construction*, where the vessels’ COG evolution was analysed to detect high variations in the COG circular standard deviation, reflecting on vessels’ change of behaviour. An example is shown in Figure 3, where one route is depicted (a), together with its analysis (b), presenting its COG evolution versus the thresholded COG circular standard deviation peaks. The figure reveals the correlation between COG and vessel behaviour changes. The complete analysis of the case of study results in the four-lane network shown in Figure 4. This figure depicts the derived maritime traffic network from AIS historical data together with an extended area enhancing the detection and accurate representation of the Traffic Separation Schemes defined within the area.



(a) **(b)**

Fig. 4. Maritime traffic network derived from the real self-reporting maritime traffic data. The synthetic representation was obtained optimising the COG threshold to 4 degrees and the COG averaging sliding window to 400 steps).

The Maritime Traffic Network Construction algorithm was built from the algorithm proposed on [12]. However, the algorithm has been optimised to allow for parallel computation considering big data challenges, and allowing distributed computing for larger geographical areas.

Maritime surveillance targets monitoring the global maritime traffic, implying the processing and storage of large amounts of information resulting in a bottleneck situation, limiting the progress of maritime situational awareness research. In this paper, an algorithm to construct maritime traffic networks has been presented, enabling efficient and compact representation of the maritime traffic using directional graphs. The historical AIS data analysed included 108,715 points composed of their coordinates, course over ground, speed over ground, vessel MMSI and timestamp. The proposed data-driven algorithm processed the data to extract only representative patterns including (1) *waypoints*, represented by their coordinates, (2) *routes*, represented by breakpoints (coordinates) and tracklets (initial and final points coordinates). The proposed schematic representation comprised 4 routes, 4 waypoints, 34 breakpoints and 34 tracklets, rounding it to a total of 106 points. Consequently, the proposed algorithm reduced the required data storage in several orders of magnitude. Finally, the knowledge discovery process calculated the a-priori probability of each vessel type navigating through each individual route and tracklet, enabling the Maritime Situational Picture to be projected on time and predicting MSP in advance for risk management.

Data compression maximisation reduces the amount of data stored with an effect on the route fitness and precision. The maritime traffic network representation accuracy depended on two parameters: (1) the COG threshold value allocated to filter out local maxima, isolating breakpoints, and (2) the averaging window size used to compute the COG circular standard deviation. The impact on the compression rate is determined by the effect of these parameters on the amount of detected breakpoints and therefore by the number of tracklets building the maritime lanes.

The impact on the network representation accuracy is estimated calculating the Root Mean Square Error (RMSE), which increases with the size of the averaging window as well as with the COG threshold. On the other hand, this leads to the decrease of the detected number of tracklets, An optimal approximation to the real route is obtained with an averaging window size between 300 and 400 steps and high COG threshold. This solution enables efficient memory storage allocation due to the small number of tracklets as well as a limited error, ranging between 6 and 8 km. This error margin should be contextualised to the area covered. Moreover, this error is directly related to the outliers included in the raw data which imply large deviations from the routes and, thus, from their compact representation, as shown in Figure 5.

3.3 Route Classification and Statistical Analysis

In this paper, we aimed not only to represent maritime traffic but also to detect areas that could be considered of higher risk in an attempt to automatically update and adjust the Maritime Situational Picture. The unsupervised construction of the maritime traffic network highlights the areas with high density traffic, pinpointing the maritime lanes as areas of interest for risk management. However, the case of study presented a complex scenario, where vessels followed different maritime lanes depending on their vessel type. More specifically, four traffic separation schemes regulated the maritime transport in the area, obliging tankers transporting pollutants or hazardous loads to follow a subroute farther

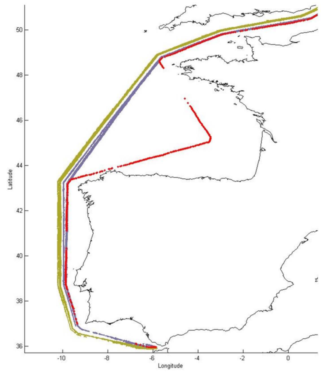


Fig. 5. Routes versus outlier representation. The figure reveals how raw data presents outliers (red) that imply large deviations from the routes and enlarge the RMSE.

from the coast, named *outer lane*. Such freight separation has a direct impact on risk management and so in the maritime situational picture. As shown in Figure 4, this separation was respected in the maritime traffic network construction. However, in order to evaluate if real-maritime traffic commits to the traffic scheme regulations, route classification was performed, matching individual vessels to the maritime lanes included in the automatically computed maritime traffic network. Moreover, such classification enabled the analysis of the vessel types navigating using each route, thus evaluating the vessels' commitment to the TSS. Figure 3.3 reveals that almost all tankers navigate using the Outer lane, which is the lane designated to tankers/cargos. However, some exceptions can be observed in Route 1, where some tankers used the Inner Lane. However, these tankers' ship type codes, 82 and 84, define their loads as medium and low pollutants. Moreover, no tankers with high pollutants used the Inner Lane. The definition of high, medium and low pollutant categories can be found in the MARPOL Annex II [3]. Regarding the other vessel types, typically the rest follow the Inner Lane as the distance traversed is shorter. Some exceptions can be observed in Figure 3.3, however, this situation has no impact on security.

4 Conclusions and Future Work

In this paper, a system to automatically derive contextual information from historical AIS data is proposed. The final objective of the system is to provide support to risk management systems by automatically updating the Maritime Situational Picture, considering transport routes as evolving entities. The presented system automatically constructs maritime traffic networks for complex scenarios, using real self-reporting data to derive contextual information, typically difficult to obtain. Different levels of granularity of the representation can be achieved by operating a trade-off between system complexity and precision.

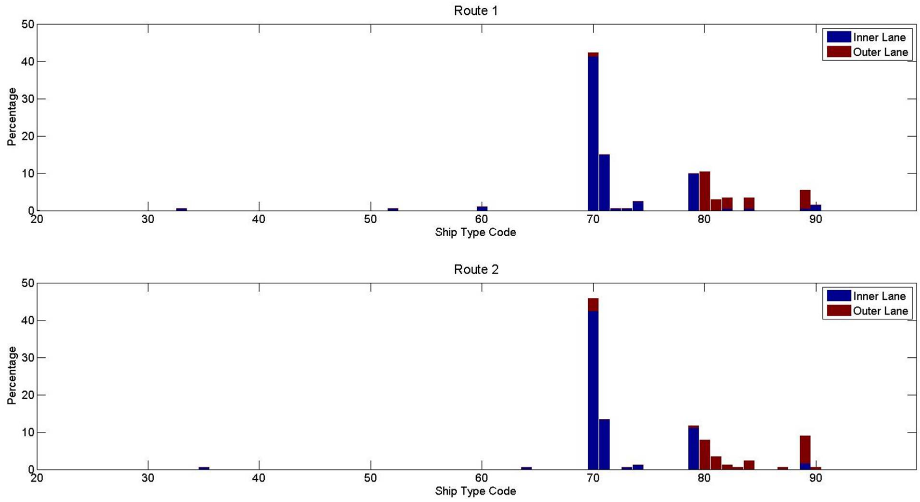


Fig. 6. Ship type presence per route. Both routes 1 and 2 are divided into four subroutes or lanes. The Outer Lane is used for tankers transporting dangerous substances, while the Inner Lane remains for the rest of the vessels which do not imply a safety, security and environment risk and so navigate closer to the coast. Such separation is regulated by the "Off Finisterre/Ushan/Cabo Roca/Cabo Sao Vicent" traffic separation schemes.

The proposed system is envisaged as the foundation for vessel position prediction, anomaly detection and risk management systems. Hence, in the future, our research will pursue three objectives. First, we will broaden the maritime network to a global scale, extending the problem to open sea areas. Second, we will address real-time anomaly detection based on real-time vessel self-reporting, implying anomaly detection based on the comparison between the maritime network and routes. Third, we will address risk management by creating risk maps, fed real-time by the maritime traffic knowledge discovery system.

Acknowledgments

The U.S. Navy and the Volpe Centre of the U.S. Department of Transport are acknowledged for the MSSIS data.

References

1. Eurostat - http://epp.eurostat.ec.europa.eu.
2. IMO - "New Amended Traffic Separation Schemes" COLREG.2/Circ.53.
3. MARPOL ANNEX II - Regulations for the control of pollution by noxious liquid substances in bulk.

4. Safety of life at sea (SOLAS) convention chapter V Regulation 19.
5. Traffic separation scheme. http://www.imo.org/.
6. Neil A Bomberger, Bradley J Rhodes, Michael Seibert, and Allen M Waxman. Associative learning of vessel motion patterns for maritime situation awareness. In *Fusion*, pages 1–8. IEEE, 2006.
7. European Commision. *Ports: an engine for growth. COM/2013/0295*, 2013.
8. European Commission. *Europe's Seaports 2030: Challenges Ahead, European Commission - MEMO/13/448*, 2013.
9. A. K. Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7, 2001.
10. M. Ester, H. Kriegel, J. Sander, M. Wimmer, and X. Xu. Incremental clustering for mining in a data warehousing environment. In *24th International Conference on Very Large Data Bases*, New York, USA, 1998.
11. M. Ester, H. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Second International Conference on Knowledge Discovery and Data Mining*, Portland, OR-US, 1996.
12. V. Fernandez Arguedas, G. Pallotta, and M. Vespe. Automatic generation of geographical networks for maritime traffic surveillance. In *Fusion*, 2014.
13. G. Pallotta, M. Vespe, and K. Bryan. Traffic knowledge discovery from ais data. In *Fusion*, Istanbul, Turkey, 2013.
14. G. Pallotta, M. Vespe, and K. Bryan. Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. *Entropy*, 15:2218–2245, 2013.
15. M. Posada, H. Greidanus, M. Alvarez, M. Vespe, T. Cokacar, and S. Falchetti. Maritime awareness for counter-piracy in the gulf of aden. In *International Geoscience and Remote Sensing Symposium*. IEEE, 2011.
16. B. Rhodes, N. A. Bomberger, and M. Zandipour. Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness. In *Fusion*, pages 1–8. IEEE, 2007.
17. Bradley J Rhodes, Neil A Bomberger, Michael Seibert, and Allen M Waxman. Maritime situation monitoring and awareness using learning mechanisms. In *IEEE Military Communications Conference*, pages 646–652. IEEE, 2005.
18. M. Riveiro, G. Falkman, and T. Ziemke. Improving maritime anomaly detection and situation awareness through interactive visualization. In *Fusion*, pages 1–8. IEEE, 2008.
19. AC Van den Broek, RM Neef, Patrick Hanckmann, Sicco Pier van Gosliga, and D Van Halsema. Improving maritime situational awareness by fusing sensor information and intelligence. In *Fusion*, pages 1–8. IEEE, 2011.

Context information in development of Geographic Information Services for inclusion of Persons with Disabilities

Ljubica Gajevic

Institute of Geoinformation, Technical University of Vienna,
Gusshausstrasse 27-29 E120-2
A-1040 Vienna, Austria
Email:ljubicagajevic@gmail.com

Abstract. This paper leads to a presentation of context information in Geographic Information Services for development of Information and Communication Technologies for free movement of pedestrians with disabilities in urban environment. It focuses on reduced mobility of persons with disability caused by inadequate physical accessibility or absence of information related to urban environment. First it explains the main characteristics of the target and basics of Design for All concepts and importance of this approach for free movement of Persons with Disabilities. Then, it shows a basic comparative analysis of needs for accessible in urban areas for blind and visually impaired people and wheelchair users since they are the most affected by obstacles and because of that the most studied groups. Finally it leads to present examples of how context information about accessibility can be used in modern Information and Communication Technologies in order to meet challenges of free movement of Persons with Disabilities in urban areas. Since there is not a lot of theoretical studies in this field this research is based on author's knowledge gained on work on specific projects for of free movement of Persons with Disabilities in urban environment. This article also presents an introduction to a PhD study.

Keywords: Persons with Disability, visually impaired, blind, wheelchair users, physical obstacles, accessibility, Design for All, context information, Geographic Information Services, Information and Communication Technologies

1 Introduction

Persons with Disabilities (PwDs) need to have access to the physical environment, transportation, information and communication technologies. The term "accessibility" is used in the Convention on the Rights of Persons with Disabilities (United Nations, 2006) as well as the term "universal design" or "Design for All" (DfA). According to the Convention (United Nations, 2006, Article 9) this approach also includes Information and Communication Technologies (ICT) as well as other facilities and services open or provided to the public, both in urban and in rural areas. Further, the

Published in the Proceedings of the 1st International Workshop on
Context-Awareness in Geographic Information Services, in conjunction with
GIScience 2014, edited by Haosheng Huang, Jürgen Hahn, Christophe Claramunt,
Tumasch Reichenbacher, CAGIS 2014, 23 September 2014, Vienna, Austria, 2014.

convention states "to ensure (…) access by Persons with Disabilities to quality mobility aids, devices, assistive technologies and forms of live assistance and intermediaries.“ (United Nations, 2006, Article 20).

Physical obstacles they face trying to access to public transport, a work place or any other service leave to a socially exclusion of this group of population. On the one hand physical obstacles in the environment should be removed and on the other hand providing specific context information for the development and use of assistive technologies should support PwDs to live independently and participate in all aspects of life even though some physical obstacles are still present. The awareness of need for DfA and taking care of Persons with Disabilities trough development of ICT is becoming more and more common. Public organizations, non-governmental organizations, and individual experts are showing rising interest in this issue.

Nowadays, online city maps are becoming increasingly popular. They can provide citizens with useful information about street names, address registers, pedestrian and bicycle routs, public transport line routes, points of interest (PoIs) and also some specific information depending on context. Number of ICTs based on Geographic Information Services (GIS) is increasing. But what can they offer to Persons with Disabilities?

2 General definitions of disability, accessibility and Design for All concept

According to the United Nations (official web page of United Nations: www.un.org) around 15 per cent of the world's population, or estimated 1 billion people, live with disabilities. They are the world's largest minority. Thus, there is an increasing awareness of social responsibility in every community to make an inclusive society, society accessible for all.

Disability according to ADA standards (ADA Standards for Accessible Design, 2010) is a generic term that includes all the components about the person: impairments, activity limitations and participation restrictions. It expresses aspects of negative interaction between the individual with health problems and physical and social environment. It means that PwDs are not people with special needs, but people with limited opportunities to explore their potential as active members of society because of inaccessible physical and social environment.

There is an increasing awareness of social responsibility in every community to make an inclusive society, society accessible for all. PwDs have to be seen as equal members of each society, not excluded, but also not specially treated.

Accessibility is a general term used to describe the degree to which a product, device, service, or environment is available to as many people as possible. Accessibility can be viewed as the "ability to access" and possible benefit of some system or entity. Accessibility is often used to focus on PwDs and their right of access to entities, often through use of assistive technology (example: The Convention on the Rights of Persons with Disabilities, United Nations, 2006).

The disability rights movement advocates equal access to social, political, and economic life which includes not only physical access but access to the same tools and services that can also improve physical access. While it is often used to describe facilities or amenities to assist people with disabilities, as in "wheelchair accessible", the term can extend to Braille signage, sign language, audio signs, etc.

Referring to equal opportunities for all the “Design for All” concept was developed. The origin of Design for All lies in the field of barrier free accessibility for Persons with Disabilities and the broader notion of universal design. The term DfA is used to describe a design philosophy targeting the use of products, services and systems by as many people as possible without the need for adaptation. DfA is design for human diversity, social inclusion and equality (EIDD Stockholm Declaration, 2004). According to the European Commission, it "encourages manufacturers and service providers to produce new technologies for everyone: technologies that are suitable for the elderly and people with disabilities, as much as the teenage techno wizards." (European Commission: Design for All, 2010) The European Commission in seeking a more user-friendly Europe and concept is about ensuring that environments, products, services and interfaces work for people of all ages and abilities in different situations and under various circumstances.

The European Concept for Accessibility Network (2003, 20 f.) also specifies some characteristics that a building environment designed for all need to fulfill. An accessible built environment has to be respectful, safe, healthy, functional, comprehensible, and aesthetic.

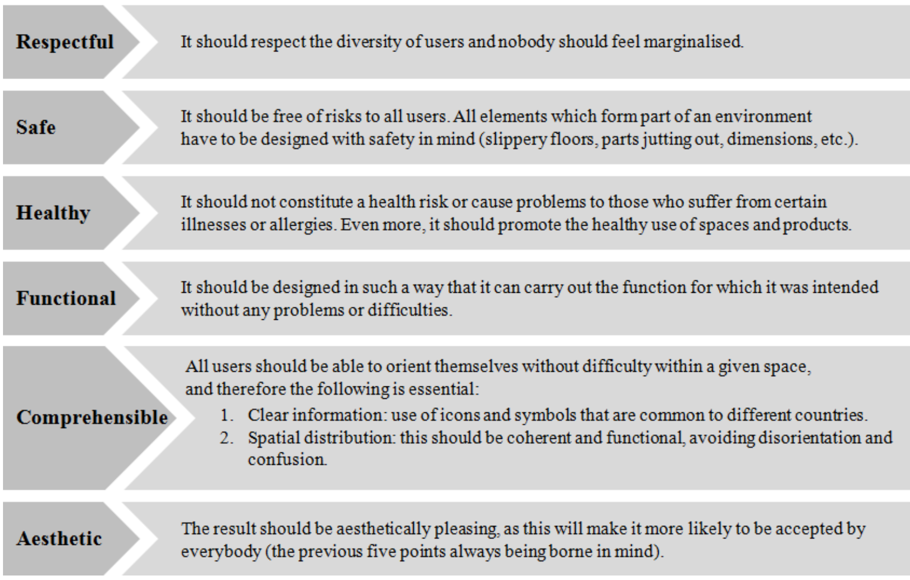


Fig. 1. Characteristics of an accessible built environment (representation based on European Concept for Accessibility Network, 2003).

This definition of the accessible building environment and DfA approach provides a solid framework for urban designers and developers. Still, in reality, environment – especially urban space – is complex and vibrant: we face fast running traffic, new objects that suddenly appear and disappear, and constructions that lead to permanent changes in cities. From one point of view this dynamic can be seen as a positive diversion from the daily monotony that is easily manageable. But have you tried to see urban space from the perspective of someone who has some kind of disability? Persons with Disabilities need right information on time in order to be able to manage these changes.

3 Context information for Persons with Disabilities seen trough obstacles and orientation points in urban environment

Obstacle in general is any physical, communication and orientation barrier that can hamper an unhindered access, movement, stay and work to persons with reduced mobility. But, when we talk about obstacles in urban space we need to keep in mind that they are different for different people. Some elements that are obstacles for one could be orientation points for the other. For example, wheelchair users are not able to access to an elevated square if there is no ramp, but for blind or visually impaired even, it is potentially dangerous, this could also be a useful marker.

3.1 Visual impaired and blind

People with severe visual impairment or blindness experience a world different from people who are sighted. When we discuss sensual perception of urban space, we have to bear in mind that on a micro-level orientation points are individual. Each route has multiple orientation points and each person perceives them in a special way. Urban space could be perceived by blind potentially either in a tactile, visual, acoustic or olfactory way. Whereas some people concentrate stronger on tactile landmarks, others recognize acoustic or olfactory ones more easily.

Obstacles for blind and visually impaired people can be summarized in:

1. Elements that cannot be recognized on time and therefore are a danger or threat for blind people because they could be hurt.
2. Elements that cannot be perceived at all, neither acoustically, visually, olfactory nor tactile.
3. Totally missing of elements.
4. Elements which are not completely reliable because they are dynamic, temporary or moving.
5. Elements which are unclear and misleading.

Complementary, non-obstacles are fixed and stable orientation points, mainly permanent installations that can be identified in a tactile way. Likewise, they are fixed acoustic guidelines and landmarks. (Neuschmid, et al., 2014)

Fig. 2. Obstacles for blind and visually impaired people: a) objects located near a tactile line that cannot be perceived by white cane; b) misleading tactile line; c) missing orientation elements (own photos)

3.2 Wheelchair users

On the other hand wheelchair users have difficulty in moving because of inadequate passage width and insuperable height differences. Their abilities are individual and depend on physical condition and type of wheelchair they use. For this group of users accessibility elements are more accurate because they are always presented in dimensions.

Generally speaking, according to accessibility standards, basic obstacles for wheelchair users are:

1. Pavements and footpaths, pedestrian crossings, parking spaces and other areas within the streets, squares, promenades, parks and playgrounds with slopes greater than 5% (in special occasions can be tolerate 8.3%) as well as width of footpaths less than 1.80 m (in special occasions can be tolerate 1.20 m) and width of free space for passage between two fixed urban elements less than 0.90 m.
2. Ramp slope greater than 5% (in special occasions can be tolerate 8.3%) with length greater than 15m and a width of less than 0.90m.
3. Inclined entrance to the plateau in front of the entrance door and with door width of less than 0.80m.
4. Public transport stops with platform for pedestrians of less than 2m in width and with height of platform that does not match the height of the floor of the wagon.
5. Public transport vehicles in which entrance is not in line with the access plateau and there is no possibility of entering a wheelchair user by ramp or other appropriate technology.

All dimensions are taken from Manual of Technical Standards of Accessibility in Serbia (Serbian Chamber of Engineers and Technicians, 2013). The manual is designed according to international standards of accessibility for PwDs.

Fig. 3. Obstacles for wheelchair users: a) pedestrian crossing with no lowered curb; b) ramp with inadequate passage width) inaccessible entrance to public transport (source: own photos)

Even the needs are individual for each person and different for both groups some similarities can be found. Basic comparative analysis of basic needs for accessible urban areas for blind and visually impaired people and wheelchair users is given in Figure 4.

	BLIND/ VISUALLY IMPAIRED	BOTH GROUPS	WHEELCHAIR USERS
Pedestrian crossings	• Acoustic traffic lights • Tactil guidelines	• Lowered curb	
Public transport stops	• Audio annuncement for vehicle arrival	• Marked entrance area • No gap beetween platform and viacle	• Annuncement for accessible vehicle arrival • Entrance in line with the access plateau and/or possibility of entering by ramp or other appropriate technology
Viacles	• Audio annuncement for each stop		
Building entrances	• Entrance is marked	• Handrail if there is a difference in height	• Entrance in line with the access plateau and/or possibility of entering by ramp or elevator, etc.
Public areas	• Tactil guidelines	• Rest areas (benches, green areas, etc.)	• Entrance in line with the access plateau and/or possibility of entering by ramp or elevator, etc.

Fig. 4. Comparative analysis of basic accessibility elements for both user groups (own representation).

4 Development of Information and Communication Technologies based on context information for free movement of Persons with Disabilities

Since in many cases Persons with Disabilities cannot rely on accessibility conditions the majority of them need to prepare their new routes in advance. Their mobility is in most of the cases highly controlled and not random at all, so a high density of information about orientation points and urban obstacles is required. Besides physical accessibility PwDs need information about accesible public transport, accessible padestrian crossings, provided acces to buildings, etc. Information about inaccessible areas are also useful for them in order to avoid them on their routs. Even though online city maps are nowadays extremely popular they are still not able to answer on needs of all user groups.

OpenStreetMap (OSM) was one of the first examples of GIS providing context data about accessibility for PwDs. OSM gives possibilities to add details on wheelchair accessibility, and the presence of things like tactile paving for blind people. OpenStreetMap provides OSM tags for all, from gluten-free food outlets, to more simple things like steps and accessible toilets. In case of tags for Persons with Disabilities the problem stands in not very detailed definition of these tags. Since OSM is an open source each user has possibility to add new tags and in most of the cases the users are not educated in accessibility field so there is a lot of mistakes as well as repeated elements.

A number of projects are delivering very interesting custom map applications based on OSM data in general as well as OSM data on accessibility. Everything started in late 1980s, when OSM still didn’t exist, but when Global Positioning System (GPS) was introduced in a civil use. First projects were developed to help visually impaired and blind. Since then there have been a huge number of projects trying to provide better accessibility for PwDs. One of the first examples was Loadstone- free satellite navigation software for blind and visually impaired people. Loadstone GPS allows a user to save points of interest and affectively create a map of the environment. Using this map a user is informed about a relative position to these points and knows the heading and distance to them. (Official web page of Loadstone: http://www.loadstone-gps.com/)

One of the current projects providing context information for blind and visually impaired people is the research project ARGUS. It is based on Galileo, the global navigation satellite system (GNSS) and data coming from Open Street Map and Citynetwork. ARGUS presents a leading climber providing a safety route for visually impaired or people working in environments with low visibility. Dangerous points that users want to avoid such as road works or potentially hazards are included. The application aims to provide the best option considering factors of safety and time. This means that the calculation of the route is influenced by black and white points on the way. Black points are potential obstacles and white points are orientation points or some affirmative spots like green areas, benches, etc. For each black point a routing algorithm prolongs the route for a number of meters in order to avoid them and for each white shortens it. Taking all this into consideration, the routing algorithm then finds the shortest way from A to B (figure 5). (Otaegui, et al., 2012)

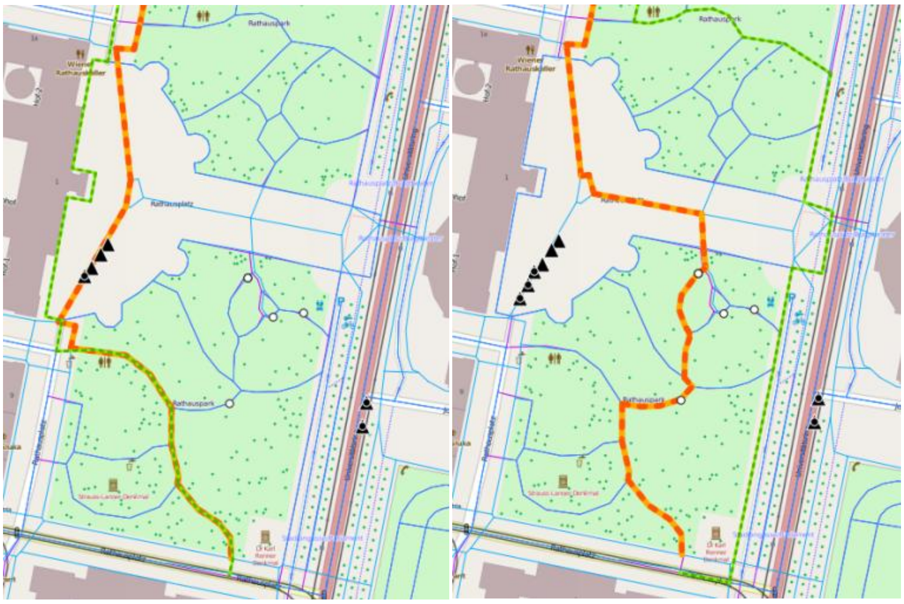


Fig. 5. ARGUS routing near the town hall in Vienna, route with black points (left) and route optimisation (right) (Source: Otaegui, et al., 2012).

Since Persons with Disabilities also face disadvantages in moving in urban areas later researchers and developers realized that a part of usage for visual impaired these applications can be used for free movement of PwDs in general, especially those in wheelchair or with reduced mobility and not only for walking distances, but also for usage of public transport. Existing information and navigation systems for public transportation mainly instruct their users in terms of transportation means, departure times and departure platforms, but do not provide them with detailed information about accessibility and accessible routs. On the one hand, missing tactile lines or path descriptions prevent blind or severe partially sighted people from maneuvering on their own at public transportation nodes. On the other hand, stairs and raised vehicle entrances are considered insuperable obstacles for wheelchair users. The number of research projects developed in this field is increasing.

5 Conclusion

As it was mentioned before Persons with Disabilities are the world's largest minority. Around 15 per cent of the world's population, or estimated 1 billion people, live with some kind of disabilities. This figure is increasing through population growth, medical advances and the ageing process, says the World Health Organization (WHO). In countries with life expectancies over 70 years, individuals spend on average about 8 years, or 11.5 per cent of their life span, living with disabilities. (Official web page of United Nations: www.un.org)

Accessibility and Design for All, these two important issues defined by Convention on the Rights of Persons with Disabilities imply design of products and services available to all people, whether they have a disability or not. Regardless, only a small percentage of PwDs is able to move independently, especially alone and on unknown routes.

Knowledge about obstacles, accessibility and DfA concept stated at the beginning is an important input for their removal as well as for planning and construction of environment accessible for all. However, this knowledge also helps in providing specific context information on accessibility and furthermore serves for development of assistive navigation and travel information systems that fulfill the requirements of all pedestrians, not only PwDs. Convention on the Rights of Persons with Disabilities do not exclude persons without disabilities so we should neither exclude them from our society.

So far several projects in the field of navigation and travel information systems have worked on this issue. With their implementation PwDs will have a greater opportunity to independently experience places and cities. However, ICT can be a supporting tool that can provide guidelines for free moving of all pedestrians. People who are not disabled, but from same reason have temporary difficulty in movement (for example travelers traveling with big suitcases or parents with small children), as well as individuals with problems in orientation or tourists can use such applications for easier movement.

No matter how different we are all have needs to move freely and independently. If we try to implement Design for All approach and design such a system that could include information relevant for all pedestrians we can make more accessible system from physical and economical point of view and also more sustainable one. We live in era of equality and it should be presented in all aspects of our life. By development of accessible transformation systems familiar with context information for all we are certainly going in a good direction.

As it was shown in a paragraph 3 there are a certain similarities in accessibility elements for blind and visually impaired people and wheelchair users. This similarity also leads to a certain similarity in context information needed. Comparative analyses of various groups of pedestrians in a domain of accessibility and context information for development of Information and Communication Technologies for pedestrians will be a topic of author's further research.

6 References

1. European Institute for Design and Disability in Stockholm, 2004. EIDD Stockholm Declaration. [online] Available at: http://www.designforalleurope.org/Design-for-All/EIDD-Documents/Stockholm-Declaration/ [Accessed 25 June 2014].

2. European Concept for Accessibility Network, 2003. European Concept for Accessibility, Technical Assistance Manual 2003. [pdf] Available at:
3. http://www.eca.lu/index.php/documents/eucan-documents/13-2003-european-concept-for-accessibility-2003/file [Accessed 28June 2014].
4. Hennig, S., Osberger, A., Neuschmid, J., Schrenk, M., Wasserburger, W. and Zobl, F., 2012. Providing Web Maps for Everyone. Understanding Users and their Requirements.In: M. Schrenk, V. Popovich, P. Elisei, ed 2012.Proceedings of 17th International CORP Conference. Schwechat, Austria, 14-16 May 2012. Schwechat: CORP, pp. 627-636.
5. MASS Factory and GABiTAP group. App&Town. Available at: http://www.appandtown.com/ [Accessed on 26 June 2014].
6. Narzt, W. and Wasserburger, W., 2013. Viator – A Mobile Travel Companion for Disabled Persons. In: M. Schrenk, V. Popovich, P. Elisei, ed 2013.Proceedings of 18th International CORP Conference. Rome, Italy, 20-23 May 2013. Schwechat: CORP.
7. Otaegui, O., et al., 2012. Argus: Assisting Personal Guidance System for People with Visual Impairment. In: Proceedings of 48th ISOCARP Congress 2012. Perm, Russia, 10-13 September 2012.
8. Serbian Chamber of Engineers and Technicians. 2013. Pravilnik o tehničkim standardima pristupačnosti [Manual of Technical Standards of Accessibility]. Official Gazette of the Republic of Serbia, no. 46/2013.
9. United Nations, 2006. Convention on the Rights of Persons with Disabilities. [online] Available at: http://www.un.org/disabilities/convention/conventionfull.shtml [Accessed on 26 June 2014].
10. US Department of Justice 2010.ADA Standards for Accessible Design. [online] Available at: http://www.ada.gov/pubs/adastatute08.htm#12102 [Accessed on 26 November 2013].
11. World Health Organization (WHO), 2014. Disability and health, Fact Sheet N°352. [online] Available at: http://www.who.int/mediacentre/factsheets/fs352/en/ [Accessed 25 June 2014].

DIA_5: An approach for modeling spatial relevancy in a Tourist Context-Aware System

Najmeh Neysani Samany[1,*], Mahmoud Reza Delavar[2], Nicholas Chrisman[3] and Mohammad Reza Malek[4]

[1]Department of GIS, Faculty of Geography, University of Tehran, Iran
nneysani@ut.ac.ir
[2]Center of Exellence in Geomatic Eng and Disaster Management, Dept. of Serveying and Geomatic Eng., College of Eng., University of Tehran, Tehran, Iran
mdelavar@ut.ac.ir
[3]Department of Geomatic Science, University of Laval, Pavillon Casault, Québec, Canada
Nicholas.Chrisman@geoide.ulaval.ca
[4]Department of GIS, Faculty of Geodesy and Geomatic Eng., K.N. Toosi Univ. of Technology, Tehran , Iran
mrmalek@kntu.ac.ir

Abstract. This article deals with a method used to manage spatial context. The proposed approach models spatial relevancy as the primary types of relevancies that determine if a context is spatially related to the user or not. The proposed approach is restricted to the urban network and assumes that in such a space, the user relates to contexts via linear spatial intervals. The main contribution of this work is that the proposed model customizes Directed Interval Algebra (DIA_{13}) to DIA_5 and applies Voronoi Continuous Range Query (VCRQ) with DIA_5 to introduce spatially relevant contexts based on the position and direction of the user. The experimental results in a scenario of tourist navigation are evaluated with respect to the accuracy and performance time of the model in 100 iterations of the algorithm on 3 different routes in Tehran. The evaluation process demonstrated the efficiency of the model in real-world applications.

Keywords: Context-awareness, spatial relevancy, customization, directed interval algebra, tourist

1 Introduction

Context-awareness is one the main topics in mobile navigation scenarios where the context of the application is dynamic. Using context-aware computing, navigation services act based on the situation of user, not only in the design process, but in real time while the device is in use (Saeedi et al., 2014). The basic idea is that location is one of the main contexts of the user and it can be modeled through spatial relevancy parameter. Context appears as a fundamental key to enable systems to filter relevant information from what is available (Dey, 2001; Schilit et al., 1994; Siewe et al., 2009;

Published in the Proceedings of the 1st International Workshop on
Context-Awareness in Geographic Information Services, in conjunction with
GIScience 2014, edited by Haosheng Huang, Jürgen Hahn, Christophe Claramunt,
Tumasch Reichenbacher, CAGIS 2014, 23 September 2014, Vienna, Austria, 2014.

Cook and Das, 2012), to choose relevant actions from a list of possibilities (Coronato and Pietro, 2011; Chedrawy and Abidi, 2006), or to determine the optimal method of information delivery (Pan et al., 2007). The major challenge of context-aware systems is the separation of the relevant from the irrelevant information (Holzmann and Ferscha, 2010; Raper et al., 2002; Reichenbacher, 2007). Among the all types of relevancies, spatial relevancy is the main types which could handle other types of relevancies such as user preferences, history, etc.(Afyouni et al., 2011). However, to the best of our knowledge, there are few reports concerning appropriate models to manage spatial relevancy parameters for a moving user in an urban traffic network.

Most of the current models for the spatial relevancy parameters in context-aware systems are based on the spatial relationships between the interacting objects (Holzmann and Ferscha, 2010; Reichenbacher, 2007). Some studies have used the proximity relations between the user and the contexts to model the spatial relevancy and utilised K-N neighbourhood or range queries (Becker and Nicklas, 2004; Neisany Samany et al., 2009). Such relations cover the inclusion of contexts in a distinct area or range and the distance to other entities (Becker and Nicklas, 2004). Holzmann and Ferscha (2010) defined a Zone-Of-Influence (ZOI) for any entity with a specified distance and direction and used the RCC5 (Cohn et al., 1997) spatial relationships to model spatial relevancy which are disjoint, overlaps, inside, contains and equal. The position, direction and extension of both ZOI are also included in their model. The most important drawback of these systems is that they do not mention the characteristic of the user's movement in an urban network which typically follows a linear route with a specific direction which doesn't have the crisp boundary (Papakonstantinou and Brujic-Okretic, 2009). Moreover these approaches do not apply the order relationships (e.g., behind or in- front- of), which could be useful in providing spatially relevant context-aware services for a moving user constrained to an urban network. Furthermore customization of spatial relations is an important fact in context-aware navigation system which has not been mentioned in our previous work (Neisany Samany et al., 2013).

The original contribution of this research with respect to our previous research (Neisany Samany et al., 2013) is the customization of Directed Interval Algebra DIA_{13} to DIA_5 which is adapted based on the position and direction of the moving user and it is able to model spatial relevancy in an urban context-aware system with Voronoi Continuous Range Query (VCRQ). Indeed the DIA_5 handles directional and topological relationships and VCRQ manages the distance in spatial relevancy model. It should be noted that customization of DIA_{13} to DIA_5 and using VCRQ in the role of Dynamic Range Neighbour Query (DRNQ) are the main ideas in this paper with respect to our previous work (Neisany Samany et al., 2013).

The proposed method is implemented in tow districts of Tehran, the capital of Iran, and we have focussed on an outdoor guided tour as an example. In this scenario, the user is a tourist who intends to visit some selected points of interest with a specified origin and destination. It is assumed that the tourist is equipped with a PDA or a laptop computer, and a GPS for positioning, and the route is constrained by a directed network. The evaluation process is based on two factors: the accuracy of the results and the time performance of the algorithm. The experimental results show that the proposed approach can effectively model and accurately detect the spatially relevant contexts within a reasonable time frame.

The rest of this paper is structured as follows: Section 2 presents fundamental aspects concerning the concepts of context-awareness and spatial relevancy, principals of modeling spatial relationships and DIA theory. The research methodology is explained in Section 3. Section 4 presents the implementation of the method and describes the experimental results with evaluating them. Finally, conclusions and directions of potential future research are considered in Section 5.

2 Background

In this section, the concept of context-awareness and spatial relevancy is explained, followed by a description of the spatial relation models with emphasis on DIA theory.

2.1 Context-awareness and Spatial Relevancy

There are different context definition in related researches: Schilit et al. (1994) explain service context as “where you are, who you are with, and what resources are nearby”; Dey (2001) defines context as “any information that can be used to characterize the situation of an entity”. In a common sense meaning, context is defined as “set of variables that may be of interest for an agent and that influence its actions” (Bolchiniet al. 2009). Context-awareness is a property of linking changes in the environment with computer systems based on relevancy of the entity (Dey 2001, Alt et al., 2009).

Saraceviec (1996) offers a general definition of relevance derived from its general qualities: “Relevance involves an interactive, dynamic establishment of a relation by inference, with intentions towards a context. Relevance may be defined as a criterion reflecting the effectiveness of exchange of information between people (or between people and objects potentially conveying information) in communication relationship, all within a context” .Various contexts in pervasive systems can be classified into primary and secondary contexts (Afyouni et al., 2011). The role of primary contexts in context management is obviously the indexing of the context information. Further information about entities can be accessed once they are found using the primary index. The identity of the entities, the location of the entities and the time are called primary contexts (Becker and Nicklas, 2004; Bettini et al., 2010). The additional context information such as user preferences, temperature, system properties and network are denoted as secondary contexts (Bonino and Corno, 2011). Following this perspective, three main relevancies in context-aware systems are identical relevancy, spatial relevancy and temporal relevancy (Becker and Nicklas, 2004). Among these relevancies, the current position – ‘the here’ – is usually the centre of action, perception and attention. Thus, the context as perceived is strongly dependent on one’s position

(Jimenez-Molina and Ko, 2011; Schmidt, 2002; Bettini et al., 2010; Holzmann and Ferscha, 2010). The identical information may be fully relevant at one position but irrelevant at another position (Bisdikian et al., 2009; Tychogiorgos and Bisdikian, 2011; Neisany Samany et al., 2013).

As seen from these observations, the locality of the context is quite important and should therefore be included in context management as one of the basic relevant parameters, called the "spatial relevancy". The spatial relevancy of an entity is dependent on the distance between the context and the user (Schmidt, 2002), their types of topological relationships and the direction of the user's movement (Holzmann et al., 2008).

2.2 Spatial Relations for Modeling Spatial Relevancy

Spatial relations are considered to be one of the most distinctive aspects of spatial information. According to Egenhofer and Franzosa (1991) spatial relations can be grouped into three different categories of topological relations, direction relations and distance relations. In this way, Allen's Directed Interval Algebra (1983) is able to consider all types of these spatial relations (Neisany Samany et al., 2013)

Being similar to the well-known Interval Algebra developed for temporal intervals (Allen, 1983); it seems useful to develop spatial interval algebra for modeling spatial relationships especially in an urban traffic network. There are several differences between spatial and temporal intervals that have to be considered when extending the intervals described by Wang et al. (2008).

There are several differences between spatial and temporal intervals which have to be considered when extending the Interval Algebra towards dealing with spatial applications (Wang et al., 2008). The most important characteristic of spatial interval is its direction. A spatial interval can have the same or the opposite direction (Renz, 2001). This leads to the definition of Directed Interval Algebra (DIA), which result from refining each relation into two sub-relations specifying either the same or opposite direction of the involved intervals, and of all possible unions of the base relations.

3 Proposed Method

The basic idea in this research is the customization of DIA in space dimension to model spatial relevancy in urban context-aware systems. This section defines the elements of the components which are adapted based on application.

3.1 Spatial Interval

The first step on using DIA is the specification of the characteristics of the directed spatial interval of the user including its extension and direction. The positive and negative directions of the directed intervals are specified as shown in Figure 1, in which the direction of the interval is determined based on its bearing of the interval (Neisany Samany et al., 2013). If the positive and negative bearing of the intervals is between 0° and 180° (0° = < bearing <180°), then the direction of the interval is positive (dirI> 0°) and if the bearing of the intervals in between 180° and 360° (180° =< bearing < 360°), the direction of the interval is negative (dirI< 0°). The extent of the spatial interval of the user is calculated using Eq. 1 (Neisany Samany et al., 2013):

$$\begin{aligned} I_{ui} = x_{cui} &= x_u - (V \times 6)\sin B_{i.i+1} \\ y_{cui} &= y_u - (V \times 6)\cos B_{i.i+1} \\ x_{cui+1} &= x_u + (V \times 6)\sin B_{i.i+1} \\ y_{cui+1} &= y_u + (V \times 6)\cos B_{i.i+1} \end{aligned} \quad (1)$$

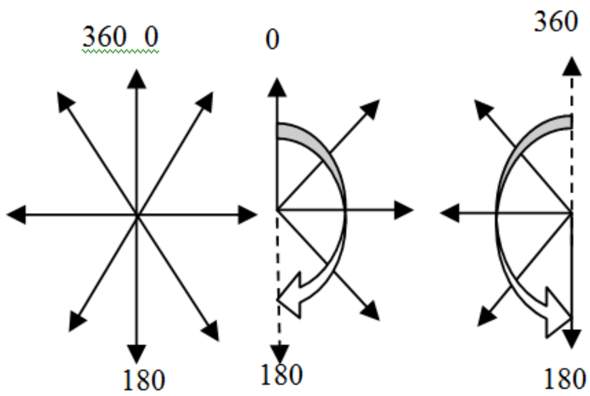


Fig. 1. Orientation relationships of the spatial directions (Neisany Samany et al., 2013)

Where I_{ui} is the moving interval of the mobile user, x_u and y_u are the coordinates of the user's position, $A_{i,i+1}$ is the bearing (B) of the directioni,i+1,(x_{cui} , y_{cui}) and (x_{cui+1},y_{cui+1}) are the coordinates of the start and end points of the directed interval respectively. As the velocity of the moving user in an urban traffic network varies, we consider V as the velocity of the user at the moment of an update and assume 6 seconds as the minimum time required by the user to make each decision during the navigation task. V×t, which is equal to distance travelled during decision making process, is used the coefficient of the bearing. Figure 2 shows a schematic view of the directed intervals (Neisany Samany et al., 2013).

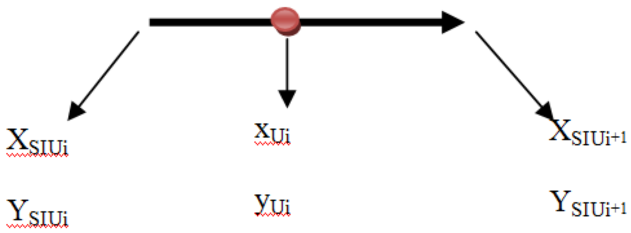


Fig. 2. A schematic view of the directional intervals (Neisany Samany et al., 2013)

The extent of urban contexts is determined with beginning and the end point of the entity as illustrated on Figure 3.

Fig. 3. The extent of urban contexts (Neisany Samany et al., 2013)

Where (x_{SIC1i} , y_{SIC1i}) is the coordinate of beginning point and (x_{SIC2i} , y_{SIC2i}) is the coordinate of end point of an urban entity along the urban traffic network.

3.2 Customizing DIA_{13} to DIA_5

Regarding to the characteristics of the moving user and related contexts in urban traffic networks, DIA could model spatial relevancy in an effective way (Neisany Samany et al., 2013). The problem is that using all of the 13 relations in spatial domain or reasoning based on DIA_{13} will reduce the speed of performance and it may decrease the efficiency of the context-aware system. Particularly when the user is moving with a specified velocity and he/she intends to make a decision due to the receiving messages of the system, the time of delivering appropriate instructions should be shortened as much as possible. Therefore it is necessary to reduce the existence spatial relationships in order to decrease computational complexity of the algorithm which leads to increase the time performance. Therefore the DIA_{13} should be customized. There are two ways for customizing the Directed Interval Algebra (Golumbic and Shamir, 1993):

To use macro relations, i.e., unions of base relations. Indeed combining IA base relations and use these macro relations as base relations is the approach of customizing.

(2) To use only the relations which is needed, namely, the interval relations <, >, d, di, o, oi, = and do not use m, mi, s, si, f, fi which correspond to intervals with common endpoints.

The main idea for customization of DIA_{13} in this paper is the combination of some of the base relations which have the same influence on decision making of the moving user to enhance the system efficiency. According to this assumption, there are five spatial relationships between the fuzzy spatial interval of the user and the fuzzy spatial interval of the related contexts including: *before (b)*, *after(a)*, *meet(m)*, *met-by (mi)* and *contact with (c)* which is combination of *overlaps*, *overlapped by*, *starts*, *started by*, *finishes*, *finished by*, *covers*, *covered by* and *equals*. We have five relations in conceptual level which is called DIA_5.

3.3 Voronio Continuous Range Query

This paper utilizes from VCRQ as *S–D continuous range search query* which is defined as: “Retrieving all objects of interest on any point during the moving of the query point from the start point (S) to the destination (D) in the networks” (Xuan et al., 2011). In the continuous environment, when the query point is moving, it will cause a series of changes on the pattern of expected searching range in respect to the moving distance of the query point during the movement. Some objects could be moved out, others could be moved in. Therefor the time of updating is determined as “every 6s”. The voronio continuous range query is carried out based on the proposed algorithm of Voronoi-based Range Search (VCRS) which is defined in (Xuan et al., 2011).

3.4 The Proposed Algorithm

The main steps of the proposed spatial relevancy model are summarised as follows (in every updating in time t when the user is moving:

Performing a Voronoi-based Range query (Xuan et al., 2011), with the centre of the user’s position and radius equal to100m. The results of this step are the preferred SI_Cs which are near to the user based on the introduced radius.

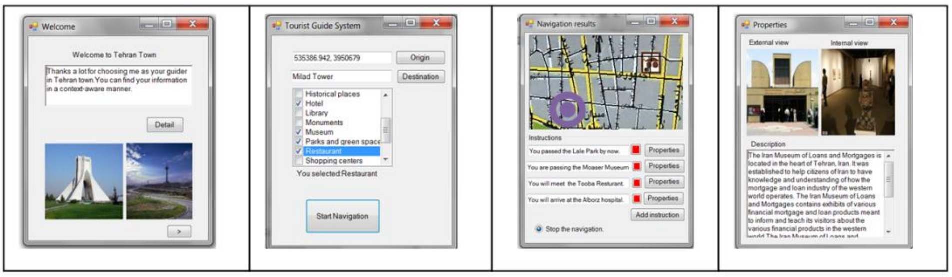
1. Definition of spatial interval of the user based on his/her position and direction
2. Specification of the spatial relationships between the DSI_U and selected SI_C based on DIA_5.
3. Sending the appropriate instruction based on detected spatial relevant contexts.

4 Implementation and Experimental Results

The proposed spatial relevancy model for an urban context-aware system is implemented in the windows application environment with the Vb.Net. programming language, using a four-stage configuration wizard as a software in the form of an exe. (or set-up) (Figure 4). The set-up file has a feature for downloading the data of the region of interest. The applied spatial data are in vector format; however, we also considered a raster map as a background (Neisany Samany et al., 2013).

4.1 Accuracy

To test this parameter, 3 independent routes were selected in the case study area and traversed in 100 iterations. In each route the related contexts selected by the tourist via the user's preference options are specified as control points. The system is run while the user moves, and the user is guided based on the spatially relevant contexts with ordered instructions. Then, the number of detected contexts in each route is compared with the control contexts. Three different metrics were used for the accuracy assessment of the proposed model including: (1) binomial approximation, (2) precision, and (3) recall.



(a) (b) (c) (d)

Fig. 4. The configuration wizard of the implemented system: a) welcome page, b) introduction of the origin and the destination to the user, specification of the preferences and the start of the navigation task, c) representation of context-aware instructions to the user and highlights of the spatially relevant contexts, d) illustration of pictures and characteristics of the selected area(Neisany Samany et al., 2013)

4.1.1 Binomial Approximation

According to the collected information, the distribution follows the form of a binomial, and therefore a General Linear Model with a one-sided binomial link function is an appropriate means to estimate the proportion detected, and a confidence interval. The results shown in Table 1 indicates the results of binomial approximation in 100 iterations of the proposed algorithm in route#1, route#2 and route#3 with 95% confidence level.

4.1.2 Precision and recall

Two other fundamental measures in the accuracy evaluation of selection process are precision and recall which are computed based on true positives (TPs, the number of relevant contexts that the system proposes to users), false positives (FPs, the contexts that have been suggested to users but they do not like), and false negatives (FNs, the contexts that have not been suggested to users but they do probably like), are (Salton and McGill, 1983):

Precision, which refers to the degree of accuracy of the selection process; it is measured as the ratio between the user-relevant contexts and the contexts presented to the user (Eq. 2):

$$\text{rcision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (2)$$

Table 1. The results of binomial approximation in route#1, route#2 and route#3

Number of entity	df	Route#1 Estimated proportion detected	Route#1 95% Confidence Interval	Route#2 Estimated proportion detected	Route#2 95% Confidence Interval	Route#3 Estimated proportion detected	Route#3 95% Confidence Interval
1	99	%99	0.973	%100	1	%100	1
2	99	%100	1	%99	0.973	%99	0.973
3	99	%99	0.973	%99	0.973	%99	0.973
4	99	%100	1	%100	1	%100	1
5	99	%98	0.956	%100	1	%98	0.956
6	99	%98	0.956	%100	1	%99	0.973
7	99	%98	0.956	%99	0.973	%98	0.956
8	99	%100	1	%100	1	%100	1
9	99	%99	0.973	%99	0.973	%98	0.956
10	99	%100	1	%100	1	%100	1
11	99	%98	0.956	%99	0.973	%98	0.956
12	99	%100	1	%99	0.973	%98	0.956
13	99	%99	0.973	%99	0.973	%99	0.973
14	99	%99	0.973	%100	1	%99	0.973
15	99	%98	0.956	%100	1	%98	0.956
16	99	%100	1	%100	1	%99	0.973
17	99	%99	0.973	%99	0.973	%99	0.973
18	99	%100	1	%100	1	%100	1
19	99	%100	1	%100	1	%100	1
20	99	%99	0.973	%99	0.973	%98	0.956

21	99	%98	0.956	%100	1	%99	0.973
22	99	%100	1	%100	1	%100	1

Recall, which is the ratio between the user-relevant contexts and the contexts present in the collection (thus also including the contexts the system does not suggest even if they can be relevant to the user) (Eq. 3):

$$\text{recall} = \frac{TP}{TP+TN} \qquad (3)$$

Table 2 shows the results of accuracy and recall parameters in 100 iterations of the algorithm in 3 different routes approach.

Table 2. The results of accuracy and recall parameters

	Percentage of Recall parameter in 100 iterations	Percentage of Accuracy parameter in 100 iterations
Route # 1	92.5	96.5
Route # 2	94.8	98.1
Route # 3	92.3	95.7

Thus, the statistics demonstrate that the proposed approach can model spatial relevancy in a context-aware system, with some degree of contexts undetected.

4.2 Time Performance of the model

In this section, the results of tests that have been performed to show the run-time efficiency of the algorithm are presented. Two performance tests were conducted, for which a Windows 7 Ultimate system (Intel® Atom (TM) CPU N270 and 2GB RAM) was used. The first one is the time with 13 fundamental spatial relations (Table 3); the second one is when he relations are reduced to 5 customized relations (Table 4).

Table 3. Time in (s) for updating the instruction when the numbers of relations are 13

The number of spatially relevant contexts in every updating	1-5	5-10	10-15
Time (second)	0.05-0.64	0.64-1.05	1.05-1.25

Table 4. Time in (s) for the instruction when the when the numbers of relations are 7

The number of spatially relevant contexts in every updating	1-5	5-10	10-15
Time (second)	0.03- 0.4	0.4- 0.71	0.71-0.85

Comparison of the proposed method with our previous work (Neisany Samany et al., 2013) demonstrated that the achieved results are improved with respect to the accuracy and time performance.

5 Conclusions and future directions

The main contribution of this paper is the specification of a model for spatial relevancy, which is adapted to the characteristics of moving user in an urban traffic network, and its implementation in a tourist guide system. The model enables context-aware services to be managed without the user's prior knowledge of the area. Adaptation of the application to the user is based on the Voronoi Continues Range Query and Directed Interval Algebra. With customizing the spatial relationships of DIA, 5 spatial relationships between intervals of the user and related contexts are specified to detect the spatial relevant contexts.

In this research the tourist guide is equipped with a PDA or Laptop system and a tool for positioning system like GPS. The tourist could execute this program in his/her device and receive the expected context-aware service conveniently. The experimental results show that the proposed approach could detect spatial relevant contexts at the right position at the right time. The right position of the context is evaluated with accuracy parameter and the right time of the context-aware services are assessed through time performance.

References

1. Afyouni, I., Ray C. and Claramunt, Ch.: Spatial Models for Indoor and Context-aware Navigation Systems: A Survey. J. Spat. Info. Sci. 4(1), 85-123 (2011).
2. Allen, J.F.: Maintaining Knowledge about Temporal Intervals. Comm. ACM 26(11): 832-843(1983).
3. Alt, F., Sahami Shirazi, A., Pfeiffer, M., Holleis, P., Schmidt, A.: TaxiMedia: An Interactive Context-Aware Entertainment and Advertising System, 2nd Pervasive Advertising Workshop, Lübeck, Germany (2009).
4. Becker, C. and Nicklas, D.: Where Do Spatial Context-models End and Where Do Ontologies Start? A Proposal of a Combined Approach. In: Proceedings of First International Workshop on Advanced Context Modelling, Reasoning and Management in conjunction with UbiComp 2004, pp. 48-53(2004).

5. Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathanf, A., Riboni, D.: A Survey of Context Modeling and Reasoning Techniques. J. PMC. 6, 161-180 (2010).
6. Bisdikian, C., Kaplan, L., Srivastava, M., Thornley, M.B., Verma, D. J., Young, R. I.: Building Principles for a Quality of Information Specification for Sensor Information. In: Proceedings of 12th Intl Conf. on Information Fusion (FUSION'09), Seattle, WA, USA, July 2009, pp. 1370 – 1377 (2009).
7. Bolchini, C., Curino C. A., Orsi, G., Quintarelli, E., Rossato, R., Schreibeer, F. A., Tanca, L.: And What Can Context do for Data?. Communications of the ACM. 52, 11, 136-140 (2009).
8. Bonino, D., Corno, F.:What would you ask to your home if it were intelligent? Exploring User Expectations about Next-generation Homes. JAISE 3, 111-126 (2011).
9. Chedrawy, Z., Abidi, S.R.: Case-based Reasoning for Information Personalization: using a Context-sensitive Compositional Case Adaptation Approach. In: Proceedings of IEEE International Conference on Engineering of Intelligent Systems, Islamabad, Pakistan, Sept. 18, pp.1-6 (2006).
10. Cook, D.J., Das, S.K.: Pervasive Computing at Scale: Transforming the State of the Art. J. PMC 8, 22–35(2012).
11. Cohn, A.G., Bennett, B., Gooday, J., Gotts, N.M.: Representing and Reasoning with Qualitative Spatial Relations about Regions. In: Spatial and Temporal Reasoning, Kluwer Academic Publishers, pp. 97-134(1997).
12. Coronato, A., Pietro, G.: Formal Specification and Verification of Ubiquitous and Pervasive System. J. ACM Transactions on Autonomous and Adaptive Systems 6 (1): 9(2011).
13. Dey, A. K.: Understanding and using Context. J. Pers. and Ubiquitous Comput., 5, 4-7(2001).
14. Egenhofer, M.J., Franzosa, R. Point-set Topological Spatial Relations. International Journal of Geographic Information Systems, 5(2): 161-174(1991).
15. Golumbic, M.C., Shamir, R.: Complexity and Algorithms for Reasoning about Time: A Graph Theoretic Approach. J. ACM 40(5): 1128–1133 (1993).
16. Holzmann, C., Ferscha, A.: A framework for utilizing qualitative spatial relations between networked embedded systems. J.PMC 6, 362-381 (2010).
17. Holzmann, C., Hechinger, M., Ferscha, A.: Relation-centric Development of Spatially-aware Applications. The 2008 IEEE 17th Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, (WETICE 2008), pp.60-65(2008).
18. Jimenez-Molina, A., Ko, I. Y.: Spontaneous Task Composition in Urban Computing Environments based on Social, Spatial and Temporal Aspects. J. EAAI 24, 1446–1460 (2011).
19. Neisany Samany, N., Delavar M.R., Saeedi S. and Aghataher R.: 3D Continuous K-NN Query for a Landmark-based Wayfinding Location-based Service. 3D Geo-Information Sciences, Lecture Notes in Geoinformation and Cartography, Part II, pp. 271-282 (2009).
20. Neisany Samany, N., Delavar, M.R., Chrisman, N., Malek, M.R.: Modelling Spatial Relevancy in Urban Context-aware Pervasive systems using dynamic range neighbor query and interval algebra, J. AISE 5(6):605-619 (2013).

21. Pan, W., Wang Z. and Gu X.: Context-based Adaptive Personalized Web Search for Improving Information Rtrieval Effectiveness. In: Proceeding of IEEE International Conference on Wireless Communications, Networking and Mobile Computing, Shanghai, China, Oct. 8-10, pp. 5427–5430 (2007).
22. Papakonstantinou, S., Brujic-Okretic,V.: Prototyping a Context-aware Framework for Pervasive Entertainment Applications, In: Proceeding of Conference in Games and Virtual Worlds for Serious Applications, pp. 84 – 91(2009).
23. Raper, J., Dykes, J., Wood, J., Mountain, D., Krause A., Rhind, D.: A Framework for Evaluating Geospatial Information. J. Info. Sci. 28(1): 39-50 (2002).
24. Reichenbacher, T.: The Concept of Relevance in Mobile Maps. Location Based Services and Tele-Cartography. Lecture Notes in Geo-information and Cartography, Section III, pp. 231-246 (2007).
25. Renz, J.: A Spatial Odyssey of the Interval Algebra: Directed Intervals. In: Proceeding of the 17th Znt 'I Joint Conference on Artificial Intelligence, B. Nebel, ed., Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, August, pp. 51–56 (2001).
26. Salton, G., McGill, M.: Introduction to Modern Information Retrieval. New York: McGraw-Hill(1984).
27. Saracevic, T.: Relevance Reconsidered. In: Proceeding of the Second Conference on Conceptions of Library and Information Science (CoLIS2), Copenhagen, Denmark, Oct. 14-17, 1996, pp.201-218 (1996).
28. Saeedi, S., Moussa, A., El-Sheimy, N.: Context-Aware Personal Navigation Using Embedded Sensor Fusion in Smartphones. Sensors, 14, 5742-5767 (2014).
29. Schilit, B.N., Adams, N., Want, R.: Context-aware Computing Applications. In: proceeding of workshop on mobile computing systems and applications, Santa Cruz, California, USA Dec. 8-9, pp. 85–90 (1994).
30. Schmidt, A.: Ubiquitous Computing – Computing in Context, PhD Thesis, Lancaster University (2002).
31. Siewe, F., Zedan, H., Cau, A.: The Calculus of Context-aware Ambient. J. Computer Syst. Sci. 77(4): 597-620 (2009).
32. Tychogiorgos, G., Bisdikian, Ch.: Selecting Relevant Sensor Providers for Meeting "your" Quality Information Needs. In: Proceeding of IEEE Conference on Mobile Data Management (MDM), Lulea, Sweden (2011).
33. Wang, Sh., Liu, D., Liu J., Wang, X.: An Algebra for Moving Objects, in: Advances in Spatio-Temporal Analysis, Taylor & Francis Group, London, , pp. 111–122 (2008).
34. Xuan, K., Zhao, G., Taniar, D., Rahayu, W., Safar, M. and Srinivasana, B.: Voronoi-based Range and Continuous Range Query Processing in Mobile Databases, J. Comp.and Syst. Sci. 77, 637–651(2011).

The Nexus of Content and Process in Context Ontology Modelling for Human Sensor Web

Habtom Tsega[1], Rob Lemmens[1] Juma Lungo[2], Menno-Jan Kraak[1], and Anna Wesselink[1]

[1] University of Twente, Department of Geo-Information Processing, Enschede AE 7500, The Netherlands,
h.b.tsega@utwente.nl,
WWW home page: http://www.itc.nl

[2] University of Dar es Salaam, Department of Computer Science, Dar es Salaam, Tanzania
WWW home page: http://www.coict.udsm.ac.tz/

Abstract. The widespread use of personal ICT devices paved the way for the devolution of geo-information to expert and non-expert users. Human sensor Web (HSW) is a notion that considers humans carrying affordable electronic and communication devices as moving sensory servers (hence 'human sensors') to generate location-oriented data. A lot of such information ends up as useful volunteered information on the web on a real-time basis.

Our initial assumption is that the efficiency of the use of citizen generated data can be improved if its context is carefully studied, modelled and incorporated as part of the information processing. The intention is to boost the smartness of the HSW system to make act upon its immediate socio-physical and system-related context. Given their wider and cross-platform applicability, ontology-based modelling techniques are our most preferred approach to represent and analyze contextual knowledge of HSW. In this paper, we attempted to deliberate and test our newly adopted approach by marrying two aspects of context modelling; (*i*) populating the *content* of the model, and (*ii*) guiding the modelling *process*.

In most cases, context ontologies are developed based on the perspectives of the developers; also known as knowledge engineers, in a top-down manner. Their main focus is the content of the ontology model, i.e., what constitues the context. A pariticipative method that admits the perspectives and feedbacks of knowledgeable domain experts is missing in the context modelling research arena. However, such methodology can be systematically borrowed from the rich experiences of the ontology engineering and software development research communities and properly married with the context modelling practices. Our work presents a combination of such approaches that complement eachother. Our approach has been tested on a mobile-based reporting system for functionality of rural water points in Tanzania in the context of SEMA (Sensor, Empowerment and Accountability) project.

Published in the Proceedings of the 1st International Workshop on Context-Awareness in Geographic Information Services, in conjunction with GIScience 2014, edited by Haosheng Huang, Jürgen Hahn, Christophe Claramunt, Tumasch Reichenbacher, CAGIS 2014, 23 September 2014, Vienna, Austria, 2014.

Author Index