Criteria of geographic relevance: an experimental study

Stefano De Sabbata* and Tumasch Reichenbacher
Department of Geography, University of Zürich, Zürich, Switzerland

PRE-PRINT VERSION

Abstract

The relevance of geographic information has become an emerging problem in geographic information science due to an enormous increase in volumes of data at high spatial, temporal, and semantic resolution, because of ever faster rates of new data capturing. At the same time, it is not clear whether the concept of relevance developed in information science and implemented for document-based information retrieval can be directly applied to this new, highly dynamic setting. In this study we analyse the criteria users apply when judging the relevance of geographic entities in a given mobile usage context. Two different experiments have been set up in order to gather users' opinions on a set of possible criteria, and their relevance judgements in a given scenario. The importance ascribed to the criteria in both experiments clearly implies that a new concept of relevance is required when dealing with geographic entities instead of digital documents. This new concept of 'Geographic Relevance' is highly dependent on personal mobility and user's activity, whose understanding may in turn be refined by the assimilation of 'Geographic Relevance' itself.

1 Introduction

In the last few years, remarkable advances in mobile computing, telecommunication services, and positioning systems have radically changed the way we create, share, and explore geographic information. Vast amounts of geographic data have become available to experts as well as to the public. The access to geographic information has also become much simpler through lean web mapping clients or applications such as Google Maps[1]. In parallel, the usage of digital geographic information has moved on to mobile platforms. The almost ubiquitous access to geographic information has become a commodity owing to affordable and powerful mobile devices, widespread availability of mobile network connectivity, provision of spatial data sets, and services built upon them, aiming at delivering location-related information to mobile users.

At the same time, rapid technological advances have led to new fundamental challenges. Mobile devices feature small displays and limited interaction capabilities. Usability studies reveal that a naïve approach to information supply is not applicable to mobile usage of geographic information (Pombinho et al., 2009). The most eminent problem arising from the large amount of spatial data available is the likelihood of human information overload. The reasons for limiting the number of geographic information objects in visual representations are: the limited capacity of information processing in both humans and mobile computers, the limitations in display space, and the need for efficiency in decision-making processes based on geographic information. The effective and efficient communication of geographic information to a mobile user in an accurate manner is a complex process. In order to address this problem we need to understand which information is relevant to a mobile user.

The issue is then how to dig into this potential mine of geographic information, in order to deliver to the users just the relevant information, and how to define relevance in this new setting. To handle this challenge, in Geographic Information Science (GIScience) the concept of Geographic Relevance (GR) has been developed (Zipf, 2003; Reichenbacher, 2005; Raper, 2007; Reichenbacher et al., 2009). GR refers

---

*Corresponding author. Email: stefano.desabbata@geo.uzh.ch
[1]http://maps.google.com

to the relevance of a geographic entity[2], given a specific context of usage. That is, GR does not refer to the relevance of a geo-referenced document or a document reporting geographic information, it refers to the relevance of the real world entity or event by itself. This definition encompasses the concept of wireless/mobile relevance proposed by Coppola et al. (2004), since it entails a situational relevance (Wilson, 1973; Saracevic, 2007) that deals with the user context and the objects in the physical world. As such, this definition is rather far away from the current understanding that underlies geographic information retrieval systems and information retrieval systems. More pragmatically, GR is intended to assess the relevance of an object, that is a representation of a geographic entity within a computer system or database. This object can be a collection of documents or an entry in a database describing a point of interest, or a combination of these two, accompanied by further information. Still, even if the evaluation is based on a single document, the objective is to approximate the relevance of the entity, not to judge how relevant the document is — in general, different documents concerning the same entity may have different levels of relevance.

Considering this new concept of GR, the succeeding issue is how to evaluate GR for a geographic entity in a given usage context. The relationship between the context of a mobile user and a geographic object in her environment is complex. It may involve not only the user's interest, but also her position, time schedule, current activity, and knowledge of the environment. Additionally, the category of the geographic object is likely to be involved, together with its location, time validity, affordance, surrounding environment, and relationships to other objects. At the moment it has not been ascertained whether these factors have an actual influence on the assessment of relevance in the setting described above, and how prominent this influence might be.

Therefore, the question is whether GR is conceptually equal to the concept of relevance employed in Information Science and Information Retrieval (IR) or not. That is, whether the same criteria used in IR can be applied to judge the relevance of a geographic entity, or if additional or different criteria — related to the factors mentioned above — have to be taken into account. The issue is then to identify an appropriate set of criteria of GR necessary and sufficient to understand the relevance of a geographic entity in a given usage context. In fact, it has been suggested that relevance criteria in IR may also be useful for GR. People would probably use the same criteria they evidently use in common IR for judging the relevance of geographic entities (Barry and Schamber, 1998). However, there are also some distinguishing criteria related to geography that have to be considered (De Sabbata, 2010). None of them has been tested yet, and others have still to be unveiled.

In this paper, we aim to report on an experimental study we conducted on the criteria of GR. In Section 2, a review of related work is presented, and the applicability of the criteria proposed in the fields of IR and Geographical Information Science (GIScience) to GR is discussed. In Section 3, four recently proposed criteria of GR (De Sabbata, 2010) are presented and a new criterion is suggested. Two human subject experiments are then presented in Sections 4 and 5. Both sections include the description of the experimental methods, the presentation of the results, and a related discussion. In Section 6, the overall results obtained in the two mentioned experiments are summarized and discussed. Section 7 concludes the paper with an outlook of future research related to Geographic Relevance.

2   Related Work

Since the dawn of IR, *topicality* has been the key criterion on which the retrieval of information has been based (Mizzaro, 1997). Generally, *topicality* is defined as the extent to which a piece of information (usually a digital document stored in an archive) concerns the topic the user is interested in. In the context of GR, the criterion *topicality* can be defined as the extent to which the category of the entity matches the user's needs in accomplishing an activity. If one is searching for a hospital, theaters are not relevant to the topic of that search. This type of category-filtering is what most Geographic Information Systems (GIS), Location Based Services (LBS), and search engines offer, together with some type of distance filtering (e.g., filtering out everything that is farther away than 300 meters from a given point).

However, '*relevance is not necessarily the same as topicality; a document on a different topic might, for one reason or another, satisfy the user's information need. Conversely, a document may not be judged satisfactory, if, for example, the patron is already familiar with its contents, or is interested in an aspect*

---

[2]As 'geographic entity', we refer to individual physical entities (or features) in the real world. GR does not aim to assess the relevance of digital documents, unlike Geographic Information Retrieval. Moreover, GR does not aim to assess the relevance, rank, evaluate the quality, or evaluate the fitness for use of entire spatial datasets.

of the topic other than that treated in the document' (Bookstein, 1979, p. 270). This led in the field of IR to investigate which criteria of relevance people use in information seeking (Boyce, 1982).

The same applies when users are searching for geographic entities or geographic information in general rather than digital documents. Then not only *topicality* and Euclidean distance — that can be considered the spatial equivalent of *topicality* — matter. This section presents the result of our search through the literature in the field of IR and GIScience, seeking for criteria of relevance that may be specifically valuable to GR.

2.1 Criteria of Relevance in IR

In the early 1990s, two different studies investigated the criteria of relevance (other than *topicality*) taken into account by the users of information systems (Schamber, 1991; Barry, 1994). Later in the same decade, the same authors jointly published a list of 10 common criteria (Barry and Schamber, 1998), such as: depth/scope/specificity, availability of information/sources of information, effectiveness, accuracy/validity, clarity, currency, tangibility, reliability/quality of sources, accessibility, and verification. This list has been the base on which subsequent work on this topic has been founded. The listed criteria are those that can be applied to almost any kind of information, and thus, they can also be applied to geographic information. In particular, the described criteria are not directly related to the properties or the attributes of a geographic entity, but they relate to how well an entity is represented in an information system, and then presented to the user.

In the following, for each criterion of the above list a definition is reported, which explains why it is valuable for GR, i.e. how it can be used in judging the relevance of a geographic entity. When referring to the source of information, the criteria refer to any person or organization from which the information about the entity that is present in the information system has been collected. These may be individuals such as the owner of a shop, or organizations such as national mapping agencies, or anything in between (e.g., web pages).

The criterion *availability* can be defined as the extent to which information or sources of information about the entity are available to the user through the information system. The criterion *specificity* (also referred to as depth or scope) can be defined as the extent to which the information about an entity has sufficient detail or depth. The criterion *accuracy* can be defined as the extent to which the information about an entity is accurate, correct or valid. The criterion *currency* can be defined as the extent to which the information about the entity is current, recent, timely, up-to-date. The criterion *verification* can be defined as the extent to which information about an entity is consistent with or supported by other information on the same subject. The criterion *reliability* can be defined as the extent to which general standards of quality or specific quality standards can be assumed, based on the source providing the information; source is reputable, trusted, expert. The criterion *accessibility* can be defined as the extent to which some effort or cost is required to obtain information (not to be mistaken with the concept of spatio-temporal accessibility in GIScience, as explained later on). The criterion *clarity* can be defined as the extent to which the information about the entity is presented in a clear and well-organized manner. The criterion *tangibility* can be defined as the extent to which the information presented to the user relates to real, tangible issues; definite, proven information is provided; hard data or actual numbers are provided. The criterion *affectiveness* can be defined as the extent to which the user exhibits an affective or emotional response to information or source of information; that is the extent to which the information or the sources of information provide the user with pleasure, enjoyment or entertainment.

Schamber (1991) identified two additional criteria that relate to how the information about an entity is presented to the user. These may have a straightforward application in a map-based system. The criterion *presentation quality* can be defined as the extent to which a source presents information in a certain format or style, or offers output in a way that is helpful, desirable, or preferable (choice of format, entertainment value). The criterion *dynamism* can be defined as the extent to which presentation of information is dynamic, active or live (e.g., presentation manipulation, zooming).

More recently, other studies have introduced further criteria of relevance, whose use can lead to an improvement of the assessment of GR. The criterion *novelty* can be defined within the context of GR as the extent to which the entity or related information are unknown or novel to the user. This criterion was first identified by Barry (1994), and then suggested by Xu and Chen (2006) as part of a five-factor model of relevance including also *topicality*, *reliability*, understandability (i.e., *clarity*), and scope (i.e., *specificity*).

Savolainen and Kari (2006) focused their work on the criteria of relevance in the context of web

searching, including image and video seeking. Part of the study was to analyse the criteria users apply when accepting (i.e., clicking on) hyperlinks in web browsing. This brought them to identify three new criteria that can be valuable for GR. The criterion *familiarity* is defined as the extent to which the user is familiar with the source of information. The criterion *variety* is defined as the extent to which the source provides a sufficient variety of information. The criterion *curiosity* is defined as the extent to which access to information is dependent on personal curiosity.

In (da Costa Pereira et al., 2009), the authors explore the criteria of personalized IR. This approach to IR strongly emphasizes the importance of the user's preferences over the popularity of web pages. The latter probably represents the main criterion applied in current web search engines. In the same study, two new criteria are proposed, which can be used in the context of GR to judge how the affordance (Jordan et al., 1998) of a geographic entity matches the current activity of the user. The criterion *appropriateness* can be defined as the extent to which the affordance of the entity is focused on the user needs. The criterion *coverage* can be defined as the extent to which the user needs are satisfied by the affordance of the entity.

2.2 Criteria of Relevance in GIScience

The two main criteria studied in the field of GIScience are spatial and temporal proximity. These criteria can be identified as part of the 'horizon' involved in the 'interpretational relevance' as defined by Saracevic (1996). The same criteria have later been encoded in the concept of 'physical relevance' by Reichenbacher (2005, 2007). As such, for this work and particularly for the experiments (see Sections 4 and 5), we are concerned with an egocentric perspective (see Raper, 2007) of representing and assessing the relevance of geographic information. That is, GR is understood in relation to the user in the centre of space and time.

The criterion *spatial proximity* can be defined as the extent to which the entity is spatially close to the user's location. A related criterion, 'geographic proximity' , was identified in a study by Schamber (1991), in which 30 users of a weather information system were interviewed. Spatial proximity is one of the main criteria used in Geographic Information Retrieval (GIR) and plays a central role in LBS.

The criterion *temporal proximity* can be defined as the extent to which an entity (or an associated event) is temporally close to the user. It may either be past, current ,or upcoming. Analyzing the criteria employed by users to evaluate the relevance of local events, Bierig and Göker (2006) observed that the usage of this criterion causes a sensible change on users' perception of the usefulness of the information.

Nevertheless, as soon as one starts taking into account the user's mobility, the two concepts of space and time can not be considered as independent anymore. In fact, users have to 'trade' time for space in order to generate mobility (Miller and Bridwell, 2009). A related concept has been proposed by Mountain and MacFarlane (2007) using the term 'temporal proximity'. The proposed concept states that entities that can be reached in a short period of time are more relevant than those that require more time to be reached. This definition does not involve just time, as the criterion proposed by Bierig and Göker (2006), but also space. In GIScience, this concept, i.e. the part of space that can be reached within some amount of time, is commonly referred to as 'accessibility'. It describes whether a user is able to interact with a geographic entity, considering the travel time the user needs to reach the entity, and the respective spatio-temporal limitations. In this study we will not use the term 'accessibility' with this meaning, in order to avoid confusion with the criterion *accessibility* as it is defined in IR (see Section 2.1). Hence, we prefer to use the term 'spatio-temporal proximity'. The criterion *spatio-temporal proximity* can be defined as the extent to which the entity (or a related event) is spatio-temporally close to the user — it may be past, current, or upcoming at the time the user will be at the location of the entity — and how long this status will last from the moment the user will have arrived at that location.

Mountain and MacFarlane (2007) also propose four filters for Mobile Information Retrieval (MIR), from which two other criteria of relevance can be derived. The first criterion is based on the assumption that a user in a mobile environment is interested in what she can see in her surroundings. The criterion *visibility* can be defined as whether the entity can be seen from the user's location or not.

Another filter developed by Mountain and MacFarlane (2007) is the 'search-ahead filter'. It is based on the assumption that users may be more interested in the entities that are on their future path, rather than those that have been passed already. This obviously implies some level of knowledge stored within the system about user's direction, future path, destination, or the usage of a prediction algorithm. That filter entails the presence of a criterion that in this paper we name 'directionality'. The criterion *directionality*

can be defined as the extent to which an entity is in the same direction the user is heading, or the amount of detour needed to include the location of the entity in the path planned by the user.

3   Criteria of geographic relevance

In the previous sections we have shown how most of the criteria studied in IR and GIScience are applicable to GR, and will probably be at the core of a GR-based system. Nevertheless, it has to be taken into account the possibility that users might consider additional criteria when evaluating the relevance of geographic entities.

In this section, we present five possible criteria of GR, that originate from well known and widely used concepts in geography (i.e., hierarchy, cluster, co-location, association rule, and anchor-point), which are novel to the concept of relevance as it has been developed in IR. Four of these five criteria have been proposed in (De Sabbata, 2010), but none of them has been evaluated yet.

If these five criteria will be found to be important for the assessment of GR, then we will have to reject the hypothesis of equivalence between GR and the concept of relevance employed in IR, which was questioned in Section 1.

3.1   The role of the geographic environment

In (De Sabbata, 2010) four new criteria of GR have been introduced. These criteria are *hierarchy, cluster, co-location,* and *association rule.* Each one of these four criteria refers to a fundamental concept in geography.

The key idea behind these criteria is that the geographic entities, which are considered in a relevance judgment, do not exist as independent entities, but rather they exist within a specific geographic context (Reichenbacher et al., 2009). These entities are commonly part of more complex phenomena which have to be taken into account when evaluating the relevance of geographic entities.

The criterion *hierarchy* can be defined as the degree of separation between the position of the user and the location of the geographic entity within a predefined spatial hierarchy. This criterion is based on the evidence that '*geographic units are cognitively and empirically organized into a nested hierarchical form (e.g., school districts)*' (Golledge, 2002, p. 8). The effect of spatial hierarchies on the judgement of distances is well documented (Stevens and Coupe, 1978). For example, users may consider an entity situated in their city district to be closer (and thus more relevant), than an entity located in a different city district (even if the actual distance is the same).

The criterion *cluster* can be defined as the degree of membership of an entity to a spatial cluster (Han et al., 2001) of related or unrelated entities. The size of the cluster can also be taken into account as a factor of relevance. We regard a cluster as a relevant area, that increases the relevance of the contained entities. Other things being equal, it can be assumed that the relevance of a single entity increases, if there are several entities of the same kind in the neighborhood (Reichenbacher, 2005).

The criterion *co-location* can be defined as the extent to which an entity satisfies a co-location pattern (Huang et al., 2004), that has been identified as common and meaningful for that category of entities within a related collection of geographic entities, and apropos the user's current needs. For example, if it is common to have restaurants close to movie theaters (since people like to go for dinner before seeing or after having seen a movie), other things being equal, a user would consider a theater with restaurants nearby to be more relevant than a theater with no restaurants nearby.

The criterion *association rule* can be defined as the extent to which an entity satisfies an association rule (Koperski and Han, 1995), that has been identified as common within a related collection of geographic entities. These rules can involve spatial, temporal, and/or other types of attribute. For example, a rule may be identified that correlates the offered services and the price range of hotels within a given zone. A user would then consider not relevant (or less relevant than others) a hotel in that zone, within that price range, that does not offer those services.

3.2   Anchor-point proximity

The concept of anchor-points is related to the notion of landmarks (Couclelis et al., 1987). There are several locations that we consider as silent clues in the environment, such as our home and work place.

Table 1: List of possible criteria of Geographic Relevance.

Properties	Geography	Information	Presentation
topicality	spatial proximity	specificity	accessibility
appropriateness	temporal proximity	availability	clarity
coverage	spatio-temporal proximity	accuracy	tangibility
novelty	directionality	currency	dynamism
	visibility	reliability	presentation quality
	anchor-point proximity	verification	
	hierarchy	affectiveness	
	cluster	curiosity	
	co-location	familiarity	
	association rules	variety	

These key locations can be considered as 'anchor' points in our understanding of the geographic environment where we live. An operational definition is not straightforward, however an anchor-point can be defined as a frequently visited location or a location where one spends a great deal of time.

In addition to the four criteria discussed in Section 3.1 we propose the criterion *anchor-point proximity*, that can be defined as the extent to which the entity is spatially close to a place that the user accounts as an anchor-point. The idea behind the criterion *anchor-point proximity* is that it is more comfortable to reach a geographic entity that is close to a well known, or frequently visited place. Given two similar places at the same distance from one's current location, if one is close to home or the work place, and the other is in a area where one is not used to go, one would probably choose the first place.

This choice can have different reasons. If one is traveling, it may be easier to reschedule one's agenda in order to reach a place near an anchor-point. If one frequently visits a given neighborhood, it may be easier to reach it and perhaps go back to that place in the future. If one knows an area, it may be easier to find an unknown place within it.

3.3 Criteria list

In Section 2, along with the key criterion *topicality*, we listed 18 criteria of relevance proposed in IR and 5 criteria of relevance proposed in GIScience. In this section, we presented 4 criteria of geographic relevance proposed in (De Sabbata, 2010), and we unveiled the new criterion *anchor-point proximity*. In total a list of 29 possible criteria of GR has been collected.

We summarize all criteria in Table 1 grouped in four classes as suggested in (De Sabbata, 2010): the class *properties* includes the criteria used in judging the entity, by means of its properties; the class *geography* includes the criteria used in judging the entity, by means of its geographical essence; the class *information* includes the criteria used in judging how well the entity is represented within the information system, by means of the available information; the class *presentation* includes the criteria used in judging how well the information is presented to the user.

4 Experiment 1

To gain a first insight into the applicability of the criteria presented in the previous sections, we designed a questionnaire to test a subset of the discussed criteria. Our main interest was to test the importance of four out of five geographic criteria presented in Section 3, that is: *hierarchy*, *cluster*, *co-location* and *anchor-point proximity*. The criterion *association rule* will be considered in further studies. A favorable feedback on those four, specifically geographic criteria would provide evidence of the difference between GR and the concept of relevance employed in IR.

4.1 Method

Participants. A total number of 132 participants took part in this experiment. A first group of 53 participants was gathered by sending e-mails to different research mailing-lists, but also groups of colleagues and friends (including researchers in Computer Science and Geography, but also non-academics). This

Table 2: Statements representing the criteria in Experiment 1.

Criterion	Questionnaire statement
Appropriateness	A place that offers just the services you need is more relevant than a place that also offers other services.
Coverage	A place that offer all the services you need is more relevant than a place that offers just some of them.
Novelty	A place that was previously unknown to you is more relevant than a place already known.
Availability	The more information available about a place, the higher is the relevance of the place.
Accuracy	The more accurate the information about a place, the higher is the relevance of the place.
Currency	The more current, recent, timely, up-to-date the information about a place, the higher is the relevance of the place.
Dynamism	The more dynamic, active or interactive the presentation of information, the higher is the relevance of the presented place.
Presentation quality	The more the information about a place is presented in a certain format or style, or offers output in a way that is helpful, desirable, or preferable, the higher is its relevance.
Spatio-temporal proximity	It is important to take into account whether the place (or a related event) will be available at the time you will be able to reach it (e.g. whether you can reach the shop before it closes).
Directionality	If you are driving, cycling, or walking, a place on your future path is more relevant than a place already passed.
Visibility	A place that is visible is more relevant than a place that you can not see from your point of view.
Hierarchy	Other things being equal (including distance), a place in the same city or district is more relevant than a place in another one.
Anchor-points proximity	A place that is close to a location you visit frequently (e.g. home or work place) is more relevant than a place in an area you are not used to visit.
Cluster	Other things being equal (including distance), a place close to a group of similar places (e.g. a shop in a shopping center) is more relevant than an isolated place.
Co-location	Other things being equal (including distance), a place that satisfies common co-location rules (e.g. a hotel with a restaurant nearby) is more relevant than a place that does not satisfy the same co-location rules (e.g. a hotel without a restaurant nearby).

first group participated in a web survey developed using the online service SurveyMonkey[3], and will be referred to as "SurveyMonkey survey" (SMs) group. A second group of 39 participants and a third group of 40 participants was gathered through the online service Amazon Mechanical Turk[4], and they will be referred to as "Amazon Mechanical Turk survey 1" (AMTs1) group and "Amazon Mechanical Turk survey 2" (AMTs2) group, respectively. We assume they fall into Amazon Mechanical Turk's demographics (Ross et al., 2010) of computers savvy people with no particular expertise in geography.

Scope. The overall idea of this first study was to ask participants about their opinions on the usefulness of the criteria we identified. The list of the 15 criteria taken into account is presented in Table 2. Not all criteria listed in Table 1 were taken into account, since our aim was to focus specifically on the geography-related criteria.

Materials. In this experiment, we used three similar web-based on-line questionnaires. As mentioned above, one was developed using the online service SurveyMonkey, and two were developed using the online service Amazon Mechanical Turk.

Procedure. The first page of each questionnaire stated the objective of the project and the purposes of the study. Then, participants were asked whether they agree or disagree (on a 7-point Likert scale) with the 15 statements presented in Table 2. Each statements represents one of the criteria taken into account.

On the second page, in the questionnaire presented to the SMs group, the participants were asked about their age and gender, and to state how frequently they use online yellow pages, digital maps, and mobile maps. On the third page, the participants were asked to rank a list of seven general criteria (summarizing the classes of criteria shown in Table 1) from the most important to the least important. Pages four and five presented the set of 15 statements (Table 2) to the participants, the first eight criteria on the fourth page and the remaining seven criteria on the fifth page. On both pages, a brief introduction was used to add some context to the questions. On the last page, an open question gave the opportunity to the respondents to specify not mentioned criteria that they would use to judge the relevance of a geographic entity, and to give us any further comment or suggestions.

In the questionnaire presented to the ATMs1 group, the 15 statements were presented to the participant at once on the second page (i.e., all the three classes in Table 2). In the questionnaire presented to the ATMs2 group, we used a slightly modified structure, that was set up in order to better fit the style commonly used in Amazon Mechanical Turk. The statements were presented to the participant on three different pages (i.e., one for each class in Table 2). In both cases, on the last page, an open question gave the opportunity to the respondents to specify not mentioned criteria that they would use to judge the relevance of a geographic entity, and to give us any further comments or suggestions.

4.2 Results

The results of this study clearly indicate that our participants agree on the usefulness of the geographic criteria. In particular, we observed a promising level of agreement on the usefulness of the four recently proposed criteria we tested (i.e., *hierarchy*, *cluster*, *co-location* and *anchor-point proximity*). These have been rated as important factors in the judgement of the geographic relevance of an entity (see Figure 1 and Table 3).

The highest rated criteria are *spatio-temporal proximity* and *coverage* with mode equal to the highest score (see Table 3). The participants also 'agree' on the importance of the criteria *currency*, *accuracy*, and *anchor-point proximity*. The majority of participants at least 'somewhat agree' (with 'agree' as the most common opinion) on the importance of a group including four geographic criteria (i.e., *co-location*, *hierarchy*, *directionality*, and *cluster*), and the criteria *availability* and *appropriateness*. Finally, lower scores have been obtained by the criteria *presentation quality* and *visibility* and even lower by the criteria *dynamism* and *novelty*. The participants seem to just 'somewhat agree' with the former, and seem to be 'neutral' with respect to the latter. The difference between these five groups of criteria is clearly illustrated by the size of the bubbles in Figure 1.

A statistically significant difference ($p < .01$) has been found for the median of the rates collected with the first questionnaire for the five criteria *availability*, *accuracy*, *dynamism*, *presentation quality* and *visibility*, with respect to the median of the rates collected with the second and the third questionnaire. That is, a statistically significant difference has been found between the responses given by the participants

---

Table 3: Mode and median values of the collected responses.

Criterion	Mode	Median
Spatio-temporal proximity Coverage	Strongly agree	Agree
Currency Accuracy Anchor-point proximity	Agree	Agree
Availability Co-location Hierarchy Directionality Cluster Appropriateness	Agree	Somewhat agree
Presentation quality Visibility	Somewhat agree	Somewhat agree
Dynamism Novelty	Neutral	Neutral

to the SMs questionnaire (mostly researchers and students in GIScience and Infromation Retrieval) and the responses collected using Amazon Mechanical Turk. This pattern can be easily spotted in Figure 1, looking at the bubbles related to these criteria. One can observe how the size of the sectors changes going from high to low values of agreement. It is important to notice that three out of five of those criteria are among the lowest rated ones. However, the origin of this difference is not clear. No statistical difference has been found between the data collected with the second and the third questionnaire. As mentioned above in Section 4.1, the detailed analysis of these differences will be presented in a further paper.

4.3 Discussion

The results presented above give us a first insight in the applicability of the single criteria of GR and a first confirmation of the importance of the geographic criteria presented in Section 3. This in turn suggests that the geographic facet of this retrieval problem appears to be really significant, and a clear indicator of a difference between GR and the concept of relevance employed in classic document-based IR.

A substantial difference between GR and classic document-based IR is also reflected by the rates of the criteria *presentation quality* and *novelty*. The first was the most mentioned criterion in Schamber's study of criteria of relevance (Schamber, 1991), and the second was the third rated in the output list of Barry (1994). The results of this survey indicate that, when a user has to judge a geographic entity rather than a document, these criteria can be accounted as secondary, maybe even optional. The same applies to the criterion *dynamism*, that can be also found in Schamber's list of criteria of relevance (Schamber, 1991). Moreover, this difference is confirmed by the agreement about the usefulness of the criteria *anchor-point proximity*, *co-location*, *hierarchy*, and *cluster*. These four new geographic criteria are a distinguishing feature of the retrieval of geographic information, and they seem to play an important role in GR.

Moreover, there is some evident discrepancy between the obtained results and the responses collected by Mountain and MacFarlane (2007) in their study on filters for Mobile Information Retrieval. In particular, the participants in our study have rated the criterion *spatio-temporal proximity* as the most important criterion, and the criterion *visibility* as one of the least important, whereas the participants in the study of Mountain and MacFarlane (2007) have rated the 'visible place' filter as slightly more desirable (those were questions about new possible filters to be developed) than the 'accessible place' filter. However, even if the criteria and the filters have been developed starting from the same concepts, the key distinct factor is the circumstance in which the concept is used. In our study, participants were asked to imagine a situation where they have to find a geographic entity in a urban environment, whereas in the study of Mountain and MacFarlane (2007) the participants were searching for information about the natural environment (e.g., tourist guide entries about plants and animals, on a mobile device) while visiting a national park. On the other hand, similar results have been obtained for the criterion *directionality*, with
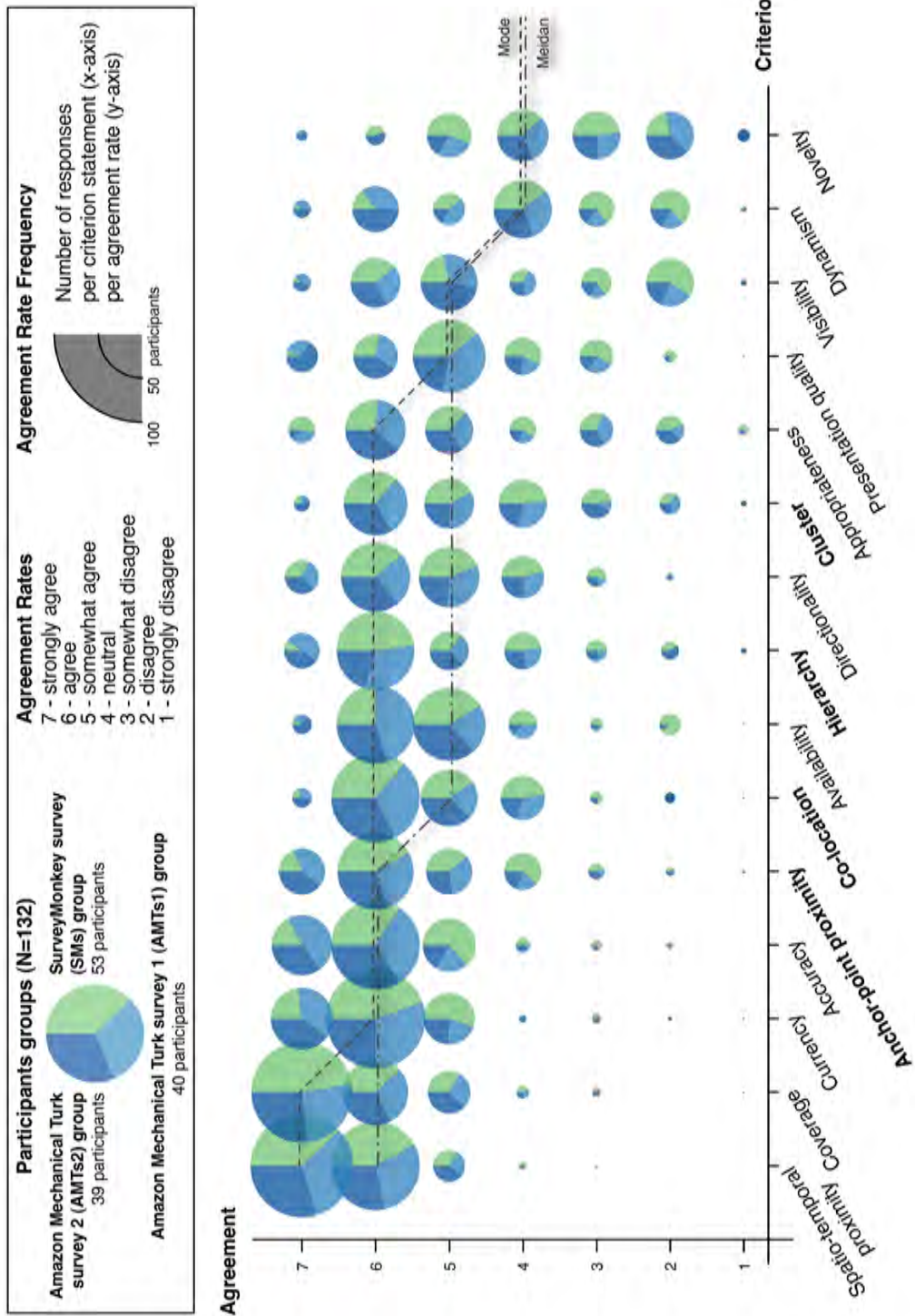
Figure 1: Summary of the responses collected from Experiment 1.

respect to the 'search ahead' filter. In summary, given the noteworthy difference between GR and the classic document-based IR, we can also assert that a certain level of difference exists between GR and MIR, even if they share various characteristics.

The study described above has two main limitations. First, there may be a difference between the answers given by a participant when asked about a criterion and the actual usage of the criterion. In fact, the role of a criterion may not be clear as soon as one has to use it in a practical situation. Second, different participants might have very different situations in mind when answering to the same question, which can influence their answers.

In the next section, a second experiment is presented where each participant is faced with an explicit mobile usage context, and given geographic information needs.

5 Experiment 2

The purpose of this experiment was to let participants directly face the geographic facet of the GR retrieval problem, in order to establish whether this fact has a significant impact when judging the relevance of geographic entities. The question was whether similar geographic entities at similar distances from the user's position would get different relevance judgements if placed in different geographic settings. In fact, most of the current geographic information systems and search engines would consider them as equally relevant, since they would have same *topicality*, *coverage* and *spatial proximity*. In particular, three of the five criteria presented in Section 3 were tested (i.e., *co-location*, *hierarchy*, and *cluster*).

The obtained results will allow us to formulate a response to the question of the equivalence between GR and the concept of relevance employed in IR, which is the main interest of our research.

5.1 Method

Participants. A total of 110 participants took part in this experiment. The participants were gathered by sending e-mails to different mailing-lists, Google Groups[5] and Yahoo Groups[6], related to the fields of IR, GIScience, and cartography. We assumed that the participants gathered by those means would have at least some familiarity with web search engines and digital maps. Each participant was randomly assigned to one of the four sub-scenarios.

Scope. The experiment was run using two different scenarios. Each scenario was composed of two sub-scenarios. Each one of the four sub-scenarios was presented to different groups of participants. In each of the two scenarios, the two sub-scenarios differed in the information presented to the participant. In particular, in the second sub-scenario, additional information was presented to the participant. This supplementary information was intended to allow the participant to apply one or more additional criteria with respect to the first sub-scenario. The objective was to compare the usage of the criteria between the different groups of participants (each one responding to a different sub-scenario).

In fact, the participants were asked to simulate (i.e., "act as", "play the role of") a hypothetical GR assessment system, taking into account all available information and the criteria they consider to be important, in order to judge the relevance of the individual geographic entities. The aim was not to test an actual application such as a geographic recommendation system or LBS.

Material. In both scenarios, the base map (see Figures 2 and 3) was derived from the geometries available on OpenStreetMaps[7] for the town of Gorizia (Friuli-Venezia Giulia, Italy), assuming that most of the participants would not be familiar with this town and thus avoid a recognition effect. The base map has been flipped vertically, the city center limits have been chosen arbitrarily (i.e., they do not reflect the actual boundaries of the town center of Gorizia), distinguishing buildings have been reshaped, and some park area have been arbitrarily added. None of the entities added to the maps (i.e., hotels, restaurants, museums, and tourist attractions) directly represent real entities in Gorizia. The three photos used in the first scenario do refer to hotels and bed&breakfasts in Gorizia, but they have been arbitrarily chosen from the images obtained by searching 'hotels Gorizia' via Google Images[8], and arbitrarily assigned to entities on the map that do not represent existing hotels in Gorizia. The reported prices and opening hours have also been arbitrarily chosen, but based on plausible values.

---

[5]http://groups.google.com
[6]http://groups.yahoo.com
[7]http://www.openstreetmap.org
[8]http://images.google.com

Table 4: Sentences representing the hypothesized criteria.

Criterion	Statement	Sub-scenarios
hierarchy	I have taken into account the distinction between the city center and the peripherial urban areas, where the first is more relevant then the others.	S1A, S1B, S2A, S2B
co-location	I have taken into account the restaurants, museums and tourist attractions, where the hotels near those POI are more relevant then the others.	S1A, S1B
availability	I have taken into account the availability of information, where the hotels presenting information about the price are more relevant then the others.	S1B
accuracy	I have taken into account the accuracy of the information about the price, where the hotels with detailed information on the price are more relevant then the others.	S1B
presentation quality	I have taken into account the quality of the presentation, where the hotels presenting an image are more relevant then the others.	S1B
spatio-temporal proximity	I have taken into account the opening hours of the restaurants, that is that the restaurants c (Today closed) and h (that will close 5 min after she could arrive there) are not relevant.	S2A, S2B
cluster	I have taken into account the groups of restaurants, where the restaurants with other restaurants nearby are more relevant then the others (e.g., if I do not find a place in one I can try in the others nearby).	S2A, S2B
directionality	I have taken into account my direction and future destination, the less I have to divert from the shortest path to the hotel, the higher the relevance of the restaurant.	S2B

Procedure. On the first page of the questionnaire, the purpose of the study was stated. On the second page, the scenario and a related map was presented (see Figures 2 and 3). The participants were asked to rate the relevance on a scale from 1 to 7 (i.e, 1 = "not relevant at all", 4 = "somewhat relevant", 7 = "extremely relevant") of a set of objects displayed on the map and to give a brief mandatory description that explained their ratings. On the third page, the participant were asked whether they used the hypothesized criteria (see Table 4). An optional comment box was provided on the third page.

Four questionnaires for the four sub-scenarios were developed using the online service OnlineUmfragen[9]. The following sections describe in detail the composition and purpose of the different scenarios and sub-scenarios.
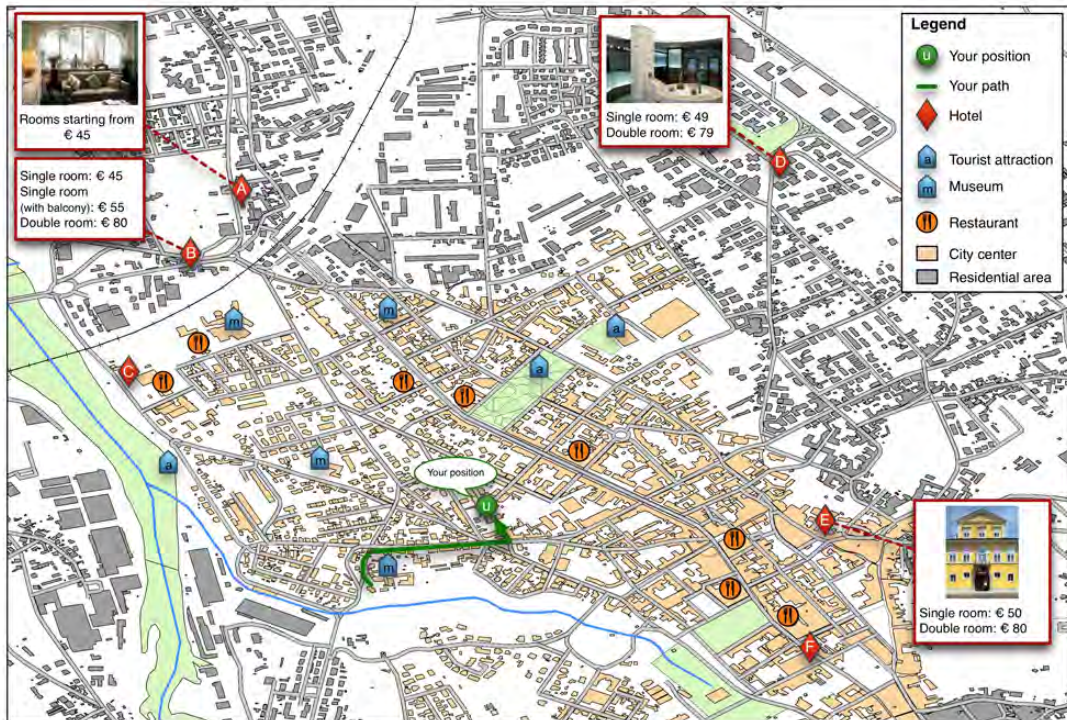
5.1.1 Scenario 1

In this scenario, participants were presented with the following situation: '*Consider the following scenario. You are visiting a city you have never been before. Just after arrival you visit one of the museums in the city center. After the museum visit you feel tired and look for a hotel for the night. Your digital city-guide on your mobile device suggests 6 hotels that fit your needs in terms of costs, availability, and offered services. The suggested hotels are all located at about the same distance from your current location. The map below indicates your current position and the 6 suggested hotels, labeled A to F. Please rate each hotel based on your needs described in the above scenario and the available information on the map*'.

Sub-Scenario 1 A A total of 28 out of 110 participants took part to the first sub-scenario (referred to as S1A). In this sub-scenario (see Figure 2(a)) the position of hotels, museums and restaurants is shown on the map, together with the position of the participant and her previous route. The city center (i.e., the touristic zone) is highlighted in a brownish color, whereas the residential areas are colored in grey.

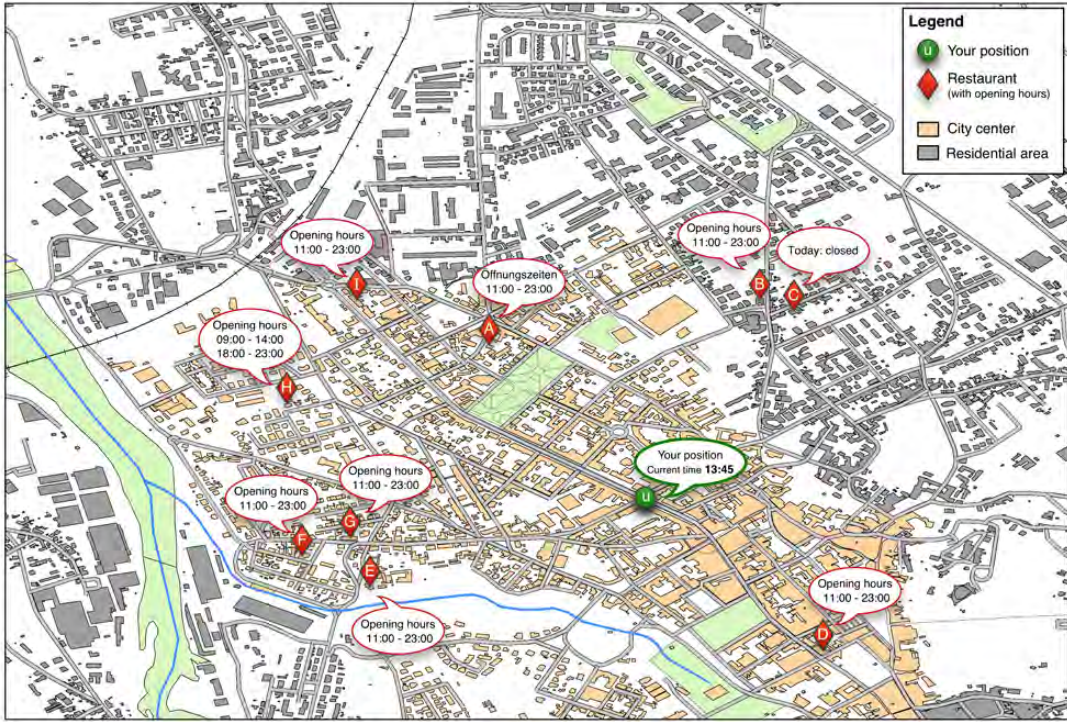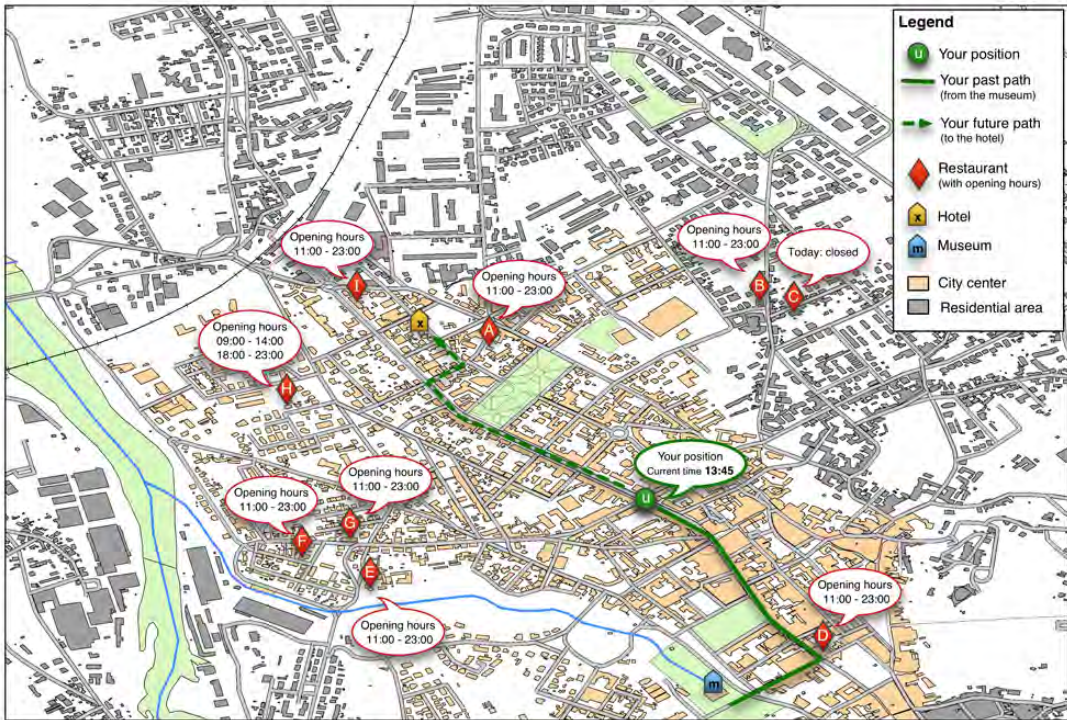[9] http://www.onlineumfragen.com
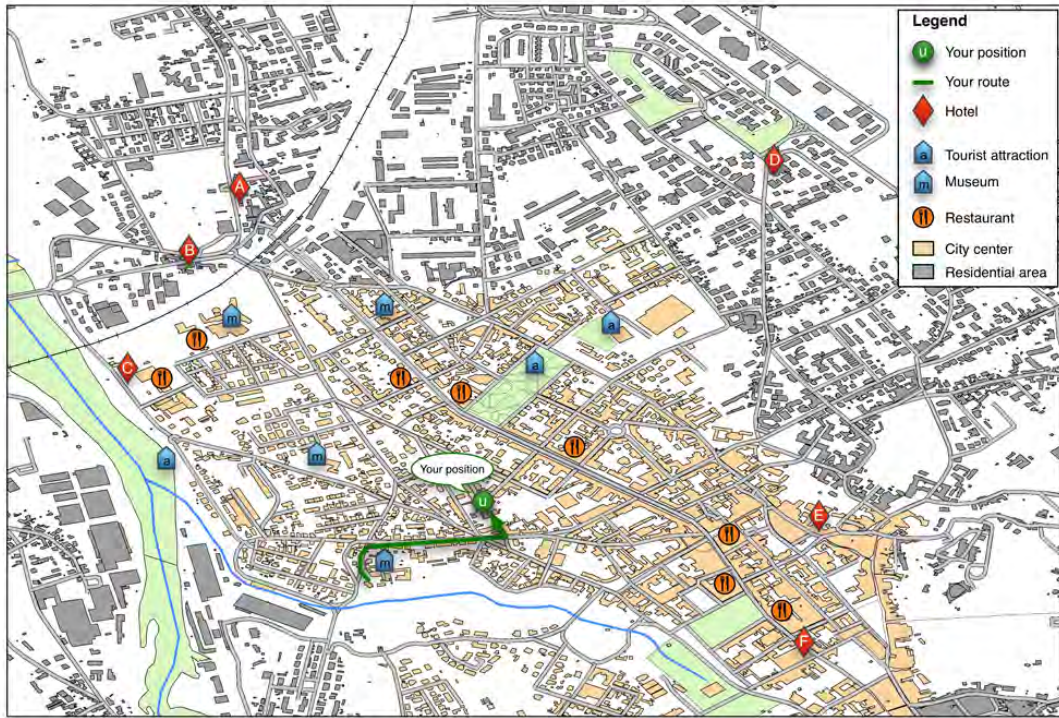
(a)

(b)

Figure 2: Scenario 1: sub-scenario S1A (a) and sub-scenario S1B (b).

(a)

(b)

Figure 3: Scenario 2: sub-scenario S2A (a) and sub-scenario S2B (b).

Three hotels are located in the city center: hotel 'C' is located near restaurants, museums and tourist attraction; hotels 'E' and 'F' are located near restaurants, with 'F' being a bit closer to them than 'E'. Three hotels are located in the residential area: hotels 'A' and 'B' are located close to the city center; hotel 'D' is located far away from the city center. We hypothesized that participants would use the available information and judge the relevance of the hotels using the criteria *hierarchy* and *co-location*, i.e.: the participant would take into account the distinction between the city center and the peripheral urban areas, where the first is more relevant than the others; the participant would take into account restaurants, museums and tourist attractions, where the hotels near those POI are more relevant than the others.

Sub-Scenario 1 B A total of 25 out of 110 participants took part to the second sub-scenario (referred to as S1B). In this sub-scenario (see Figure 2(b)) the position of museums and restaurants is shown on the map, together with the position of the participant and her previous route. The position of hotels is also displayed, and in some cases it is accompanied by some further information on the price of the room or a hotel picture. The hotels are placed in the same position as they were placed in sub-scenario S1A. Detailed price information and a picture have been attached to hotel 'E' (located in the city center, quite close to some restaurants) and to hotel 'D' (located in the residential area, far away from the city center). Detailed price information has been attached to hotel 'B' and a picture has been attached to hotel 'A', which are close to each other, just outside of the city center. No further information has been attached to the remaining two hotels. We hypothesized that participants would use the available information and judge the relevance of the hotels using the two criteria mentioned in the first sub-scenario (S1A), along with the criteria *accuracy*, *availability*, and *presentation quality*, i.e.: the participant would take into account the accuracy of the information about the price, where hotels with detailed price information are more relevant than the others; the participant would take into account the availability of information, where the hotels presenting information about the price are more relevant than the others; the participant would take into account the quality of the presentation, where the hotels presenting an image (that is, presenting information about the hotel in a way that is common to be found in touring guides and websites) are more relevant than the others.

5.1.2   Scenario 2

In this scenario, participants were presented with the following situation: '*You are visiting a city you have never been before. Just after arrival in the early morning you visit one of the museums in the city center. At 13:45H you are hungry and decide to have a late lunch. Your digital city-guide on your mobile device suggests 9 possible restaurants that fit your needs in terms of cost and offered dishes. The suggested restaurants are all located at about the same distance from your current location: a 10 minute walk. You have not booked a table at any of those restaurants and you do not know anything about table availability either. The map below indicates your current position and the 9 suggested restaurants, labeled A to I, including their opening hours. Please rate each restaurant based on your needs described in the above scenario and the available information on the map*'.

Sub-Scenario 2 A A total of 28 out of 110 participants took part to the first sub-scenario (referred to as S2A). In this sub-scenario (see Figure 3(a)) the position and opening hours of the restaurants are shown on the map, together with the position of the participant and the current time. The city center (i.e., the touristic zone) is highlighted in a brownish color, whereas the residential area are colored in grey. Seven restaurants are located in the city center. Three of them have been placed in order to form a cluster (these are the restaurants 'E', 'F', and 'G'), the other four restaurants have been placed in function of their role in the second sub-scenario (as explained in the next paragraph). Two more restaurants are located in the residential area, close to each other. We hypothesized that participants would use the available information and judge the relevance of the restaurants using the criteria *spatio-temporal proximity*, *hierarchy*, and *cluster*, i.e.: the participant would take into account the opening hours of the restaurants, that is that the restaurants 'C' (today closed) and 'H' (will close 5 minutes after she could arrive there) would not be relevant; the participant would take into account the visible distinction between the city center and the peripheral urban areas, where the first is more relevant than the others; the participant would take into account the visible clusters, where the restaurants that are part of a cluster would be more relevant than the others (if she does not find a place in one she can try in the others nearby).

Table 5: Differences in the rates of the hotel between S1A and S1B. None of the hotels had normally distributed rates. In order to double-check the robustness of the results of Mann-Whitney U tests with respect to the number of participants, T-tests have been run on the same data, then the power of the T-tests with significant results have been calculated.

Hotels	A	B	C	D	E	F
Mann-Whitney U	226.0	267.0	232.0	220.0	264.0	208.0
p	$< .05$	$> .05$	$< .05$	$< .05$	$> .05$	$< .05$
T-test F	0.12	0.51	0.69	2.40	0.44	0.19
p	$< .05$	$> .05$	$< .05$	$< .05$	$> .05$	$< .05$
Power	.70		.67	.68		.82

Table 6: Differences in the rates of the restaurants between S2A and S2B. None of the restaurants had normally distributed rates. In order to double-check the robustness of the results of Mann-Whitney U tests with respect to the number of participants, T-tests have been run on the same data, then the power of the T-tests with significant results have been calculated.

Restaurants	A	B	C	D	E	F	G	H	I
Mann-Whitney U	107.0	322.5	362.5	185.0	172.5	163.0	162.5	383.0	122.5
p	$< .01$	$> .05$	$> .05$	$< .01$	$> .01$	$< .01$	$< .01$	$> .05$	$< .01$
T-test F	87.28	.85	10.26	5.75	.00	.00	.00	1.81	10.84
p	$< .01$	$> .05$	$> .05$	$< .01$	$> .01$	$< .01$	$< .01$	$> .05$	$< .01$
Power	.99			.99	.98	.98	.98		.99

Sub-Scenario 2 B A total of 29 out of 110 participants took part to the second sub-scenario (referred to as S2B). In this sub-scenario (see Figure 3(b)) the position and opening hours of the restaurants are shown on the map. The participant knows the path she has cone from the museum to her current location (which is displayed on the map), and she knows that, after lunch, she has to go back to the hotel (which is displayed on the map) in order to be able to pack her stuff and leave in the early afternoon. The restaurants are located in the same position as they were placed in sub-scenario S2A. Restaurant 'A' is located very close to the hotel and very close the participant's planned path, whereas restaurant 'I' is located at the same distance, but on the other side. Restaurant 'D' is located on the participant's past path, in the opposite direction with respect to the participant's future path. We hypothesizes that the participant would use the available information and judge the relevance of the hotels using the three criteria mentioned in the first sub-scenario, along with the criterion *directionality*, i.e.: the participant would take into account her direction and future destination — the less one has to divert from the shortest path to the hotel, the higher the relevance of the restaurant.

5.2 Results

Looking at the responses collected for sub-scenario S1A (see Figure 4(a)), we observe that hotels 'C' ($\mu_{1/2} = 6.0$) and 'F' ($\mu_{1/2} = 6.0$) have been rated as relevant by the participants, whereas hotel 'D' ($\mu_{1/2} = 1.0$) has been rated as not relevant. The highly rated hotels are both in the city center and near by restaurants or museums. On the contrary, the low rated hotel is in a residential area and far away from any restaurant or museum. The ratings assigned to the three remaining hotels (i.e., 'A', 'B', and 'E') occur with equal probability, so their relevance is undetermined. The criteria *hierarchy* and *co-location* have been widely cited in the explanations given by the participants (7 and 19 mentions out of 28 participants) and widely used according to the answers on the third page (78.6% and 85.7%).

Looking at the responses collected for sub-scenario S1B (see Figure 4(b)), we observe that hotel 'E' ($\mu_{1/2} = 6.0$) has been rated as relevant by the participants, whereas the hotel 'B' ($\mu_{1/2} = 4.0$) has been rated as somewhat relevant. The highly rated hotel is in the city center, near by some restaurants, and displaying exact prices of the rooms together with a picture. The middle rated hotel is in a residential area, but not too far from the city center area (and thus from restaurants and museums); the detailed prices of the rooms of this hotel are displayed. The ratings assigned to the four remaining hotels (i.e., 'A', 'C', 'D', and 'F') occur with equal probability, so their relevance is undetermined. These are also
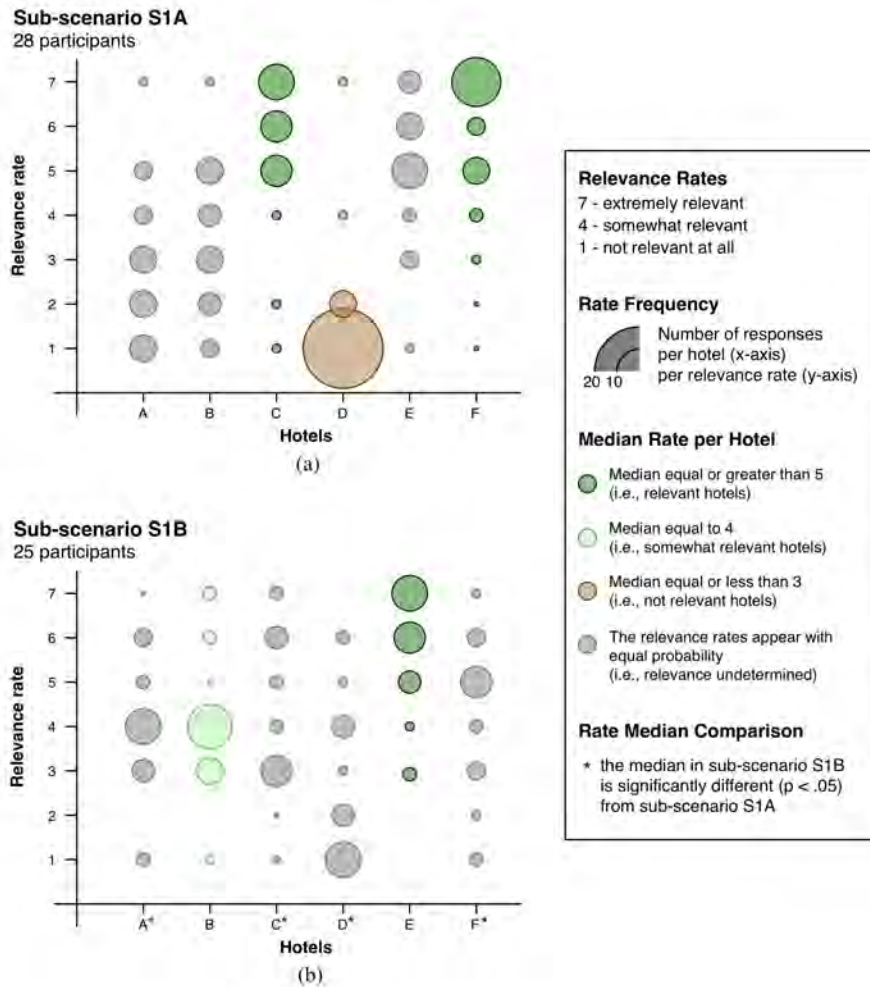
Figure 4: Results obtained for Scenario 1: sub-scenario S1A (a) and sub-scenario S1B (b).

the same hotels whose rates are significantly different (see Table 5) in the median with respect to the sub-scenario S1A. The criteria *hierarchy* and *co-location* have been widely cited in the explanations given by the participants (7 and 13 mentions out of 25 participants) and widely used (65.4% and 57.7%). The criterion *availability* has also been well cited in the explanations (6 mentions out of 28 participants) and used (50.0%). The criteria *accuracy* and *presentation quality* have been far less cited (1 and 8 mentions out of 28 participants) and used (23.1% and 30.8%).

Looking at the responses collected for sub-scenario S2A (see Figure 5(a)), we observe that restaurants 'E' ($\mu_{1/2} = 7.0$), 'F' ($\mu_{1/2} = 7.0$), and 'G' ($\mu_{1/2} = 7.0$) have been rated as relevant by the participants, whereas restaurants 'C' ($\mu_{1/2} = 1.0$) and 'H' ($\mu_{1/2} = 1.0$) have been rated as not relevant. The highly rated restaurants are the ones of the cluster in the city center, which are also open full-time today. On the contrary, the low rated restaurants are the one that is 'closed today' and the one that is going to close in 5 minutes after the participant's arrival in loco. The ratings assigned to the four remaining restaurants (i.e., 'A', 'B', 'D', and 'I') occur with equal probability, so their relevance is undetermined. The criteria *spatio-temporal proximity* and *cluster* have been widely cited in the explanations given by the participants (17 and 16 mentions out of 28 participants) and widely used (89.2% and 78.6%). The criterion *hierarchy* has been less cited (8 mentions out of 28 participants) and used (42.8%).

Looking at the responses collected for sub-scenario S2B (see Figure 5(b)), we observed that restaurants 'A' ($\mu_{1/2} = 7.0$) and 'I' ($\mu_{1/2} = 6.0$) have been rated as relevant by the participants, whereas restaurants 'B' ($\mu_{1/2} = 2.0$), 'C' ($\mu_{1/2} = 1.0$), 'D' ($\mu_{1/2} = 2.0$), and 'H' ($\mu_{1/2} = 1.0$) have been rated as not relevant. The highly rated restaurants are the ones nearby the participant's hotel and in the direction of her planned
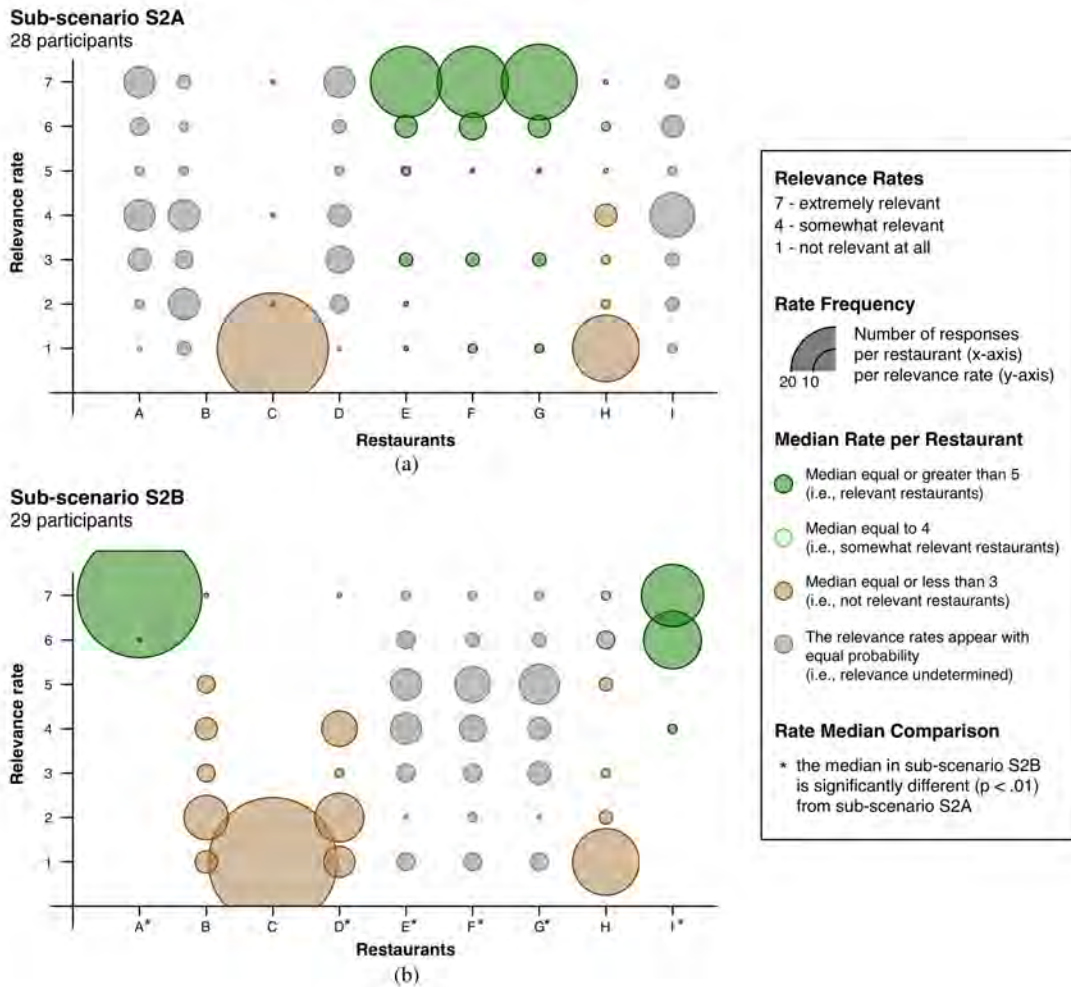
17

Figure 5: Results obtained for Scenario 2: sub-scenario S2A (a) and sub-scenario S2B (b).

path, which are also open full-time today. On the contrary, the low rated restaurants are the one that is 'close today', the one that is going to close in 5 minutes after the participant's arrival in loco, the one she passed by coming from the museum to her current location (that is in the opposite direction with respect to the hotel), and the one out of the city center. The ratings assigned to the three remaining restaurants (i.e., 'E', 'F', and 'G') occur with equal probability, so their relevance is undetermined. Almost all restaurants obtained rates significantly different (see Table 6) from the ones obtained in the sub-scenario S2A, apart from the two restaurants with very low rates in both sub-scenaios (i.e., 'C' and 'H') and restaurant 'B'. The criterion *directionality* has been cited in 26 out of 29 explanations and used by all but one of the participants according to the answers on the third page (96.6%). The criteria *spatio-temporal proximity* and *cluster* have been widely cited in the explanations given by the participants (12 and 7 mentions out of 29 participants) and widely used (75.9% and 51.8%). The criterion *hierarchy* has been used just by a small number of participants (17%).

5.3 Discussion

In general, we find that similar geographic entities at similar distance from a user's position do get different relevance judgements, if placed in different geographic settings. The responses collected in this experiment confirm the insights gained from the first experiment (see Section 4), and suggest a rejection of the hypothesis of equivalence between GR and the concept of relevance employed in IR. The results also confirm the importance of the three geographic criteria tested, and reassert the uncertainty about

other criteria.

The importance of the criteria *co-location* and *cluster* clearly emerges from the results and is supported by comments obtained from the first and second scenario respectively. The closeness of a hotel to points of interest, such as restaurants and museums, seems to be a good criterion to identify highly relevant hotels. Still, the co-location of hotels with restaurants seems to play a more important role than the co-location of hotels with tourist attractions and museums. In fact, among the meaningful co-location rules that can be taken into account when computing GR, different co-colocation rules may have different importance. Concerning the criterion *cluster*, the entities in the second scenario that were part of a cluster (i.e., 'E', 'F', and 'G') do obtain higher rates specifically because they are part of that cluster, as stated in the explanations provided by the participants about their ratings. However, it is interesting to note how none of the participants considered the entities 'B' and 'C' in the second scenario as a cluster. This may be due to the small number of entities (i.e., two is not enough to form a cluster), but it is probably largely influenced by the fact that 'C' was not spatio-temporally available.

The importance of the criterion *hierarchy* is evident in the first scenario (e.g., compare the different ratings of 'D' and 'E'), given the explanations provided by the participants about their ratings. Nevertheless, in both the sub-scenarios S1A and S1B, those entities which are located in the residential area are also farther away from tourist attractions and restaurants. That is, the criteria *co-location* and *hierarchy* are not fully independent factors in this case. Moreover, in sub-scenario S2A, one third of the participants took into account this criterion, but decided not to use it, and in sub-scenario S2B, two thirds of the participants did not think about this criterion. The criterion *hierarchy* is somewhat important, but it seems to be superfluous in the second scenario, where the use of the other criteria seems to be enough to make a decision about the relevance of an entity.

The outstanding importance of the criterion *spatio-temporal proximity* becomes evident by the entities 'C' and 'H', that have been clearly rated as not-relevant in the second scenario. The difference between the rates obtained in sub-scenario S2A and the rates obtained in sub-scenario S2B for all entities (except 'C' and 'H', which are not relevant in both sub-scenarios) is a prominent evidence of the importance of the criterion *directionality*. It is also important to notice that this criterion has been mentioned in the explanation of almost all participants, but with some differences. The responses can be categorized in three different connotations of the concept, that is: the 'easiness' of route to a final destination (i.e., the hotel in sub-scenario S2B) when including the additional stop; the length (to be minimized) of the deviation from the planned route to the final destination; and the proximity to the final destination. In fact the first two are very similar, whereas the third is notably different. Most participants in sub-scenario S2B mentioned the third connotation, whereas the first two have been mentioned just among those respondents who rated 'A' as more relevant than 'I'.

The importance of the criterion *availability* is also very clear. In sub-scenario S1A (see Figures 2(a) and 4(a)), the responses entail 'F' as slightly more relevant than 'E', because 'F' was closer to the restaurants and to the main road than 'E', given the explanations provided by the participants about their ratings.. In sub-scenario S1B (see Figures 2(b) and 4(b)), the availability of further information marks a clear distinction between the two entities: 'E' (which is displayed together with price information and an image) have been rated as more relevant than 'F' (which has neither price information nor image) by 13 out of 25 participants. Nine participants have assigned the same level of relevance both objects, and 'F' has been rated more relevant than 'E' by just 1 participant.

The role of the criteria *accuracy* and *presentation quality* is rather unclear. Roughly a quarter of the participants have used the first criterion, one third of the participant stated to have used the second criterion, and one third of the participants would use them, but did not think about it. The entities 'A' and 'B' in the first scenario were very close to each other, in the same type of area, and at the same distance from the points of interest. In sub-scenario S1B, the entity 'A' was shown with vague information about the room price and an image, the entity 'B' was shown with precise information about the room price but with no image. From the collected answers, 5 out of 25 participants have considered 'A' as more relevant than 'B' (4 of them have directly mentioned the presence of a picture as motivation), 6 participants have considered 'B' as more relevant than 'A' (but just one of them have mentioned the accuracy of the information as a motivation), and 14 participants have given the same rate to the two entities. Nevertheless, hotel 'B' has been altogether rated as somewhat relevant, whereas the overall relevance of hotel 'A' is undetermined. This is consistent with the result of the first experiment, where the criterion *accuracy* had been indicated as more important than the criterion *presentation quality*. It may be that the importance of these criteria is more related to personal preferences than the previous criteria. It may also be that the used example was not appropriate to understand the relevance of these

criteria, which can be more important in other specific situations.

In sub-scenario S1B, it is also difficult to unquestionably distinguish the influences of the criteria *accuracy* and *presentation quality* from the influence of the criterion *availability* (which appear to be stronger, as explained above), and these three criteria are not fully independent from one another. Still, in general, the effect of these criteria on the overall relevance of the entity seems to be rather narrow. In fact, hotel 'D' in sub-scenario S1B has been rated as the least relevant among the entities of that sub-scenario — as it was in S1A, where no further information was given — even if it was presented with detailed price information and a picture.

6   General discussion

Given the results obtained from the two experiments described above, we reject the hypothesis of equivalence between GR and the concept of relevance employed in IR. Therefore, we argue that GR and the concept of relevance employed in IR are different, because of the geographic and mobile facets that concerns the first concept but not the second. It is clear that the proposed criteria of *co-location*, *cluster*, and *hierarchy* play an important role in the judgment of the relevance of geographic entities, and that these criteria are a distinguishing feature of GR with respect to the classic document-based IR. The criterion *anchor-point proximity* has not been fully tested, but the results obtained from the first experiment suggest that it is an important criterion of GR.

The objective of this study was not to provide a finite, stable, ordinate set of criteria for a specific application of GR. We also do not mean to state the recently proposed criteria as the most important criteria in the retrieval of geographic information. Nevertheless, aiming to understand the importance of the five new geographic criteria, we faced the problem of understanding the concept of 'importance of a criterion', which seems to be as fuzzy as the concept of 'relevance' itself.

There are some criteria, such as *topicality* or *spatio-temporal proximity*, that define whether an entity is relevant or not. If one is interested in hospitals, theaters are not relevant; if an entity is not temporally available in the time range one needs it, it is not relevant. These few 'fundamental' criteria can be used to filter out options that do not fit at all the user's needs in terms of user's interest or mobility limitations (Miller and Bridwell, 2009).

A second set of criteria is composed by those that define how much relevant a feasible option is. These can be labeled as 'primary' criteria. This group includes those criteria which in our second experiment implied a significant difference in the rates, such as *directionality*. We argue that the criteria *co-location*, *cluster* and *hierarchy* are part of this group. This argument is supported by results obtained in both experiments, in particular: the results obtained for the first scenario in the second experiment regarding the criteria *co-location* and *hierarchy*, and the results obtained for the second scenario in the second experiment regarding the criterion *cluster*.

A last, third set of criteria is composed by those that can help to distinguish between two similar entities, which are however non-compensatory criteria — that is, they do not have a significant impact on the relevance of an entity. These criteria can be labeled as 'secondary' criteria. For example, in the sub-scenario S1B in the second experiment (see Section 5.1.1), the better *presentation quality* of hotel 'A' does not compensate for the lack of information *accuracy*. It follows that hotel 'A' has been overall rated as less relevant than hotel 'B', which has lower *presentation quality* but higher information *accuracy* — given an undoubtedly equal relevance in the matter of the criteria *availability*, *hierarchy*, and *co-location*.

The importance of the criteria in the third set can be related to the context of the search (e.g., reference images may be quite important if one searches for a hotel, probably less important if one searches for a post office) or personal preferences. This group includes criteria such as *presentation quality* and *visibility*, where the importance of the former is probably dependent on personal preferences and the importance of the later may be dependent on the situation (e.g., more important in open spaces than in an urban environment). Further studies are needed to better understand how the importance of a criterion can be related to the situation the criterion is used in.

A further interesting point unveiled in our second experiment concerns the 'Geography' set of criteria (see Table 1). Analysing the explanations given by the participants to justify their responses, it is evident that this group is not homogeneous, but it is composed of two distinct sub-groups of criteria. These are criteria that are commonly mentioned together in the explanations, sometimes they are combined in some more general concept. A first group is related to the user's personal mobility, and includes criteria such as *spatio-temporal proximity* and *directionality*. These criteria are 'dynamic', that is, the relevance of an

entity with respect to these criteria can be calculated only in a given situation (e.g. user's position, time schedule and mode of transportation) and changes as the situation changes. A second group is related to the geography of the environment, and includes criteria such as *co-location* and *cluster*. These criteria are more 'static', that is, the relevance of an entity with respect to these criteria can be calculated just once 'off-line', independently from the given situation in which the user will apply these criteria. Clearly, these are not static in every respect, changes in the environment may imply changes in the relevance of the entities. Still, it seems conceptually difficult to apply this classification to the criterion *anchor-point proximity*. This criterion relates more to the user than to the geographic environment, but it is 'static' at the same time.

The user's personal mobility and the geography of the environment are the two discernible geographic aspects of GR. Both aspects have a great influence on the relevance of the geographic entities. It is well known how those two factors also influence the user's activity (Raubal et al., 2004). It is also well known that the user's activity influences the user's mobility (Miller and Bridwell, 2009) and the relevance of the geographic entities (Reichenbacher, 2009). Thus, a topic to be further investigated is if (and how) the relevance of the geographic entities influences the user's mobility and activity, and then, how the assessment of GR can modify and improve our understanding of personal mobility, activities, and their relations to each other.

7 Conclusions

In this paper, we presented a list of 29 criteria, which concern the geographic relevance of an entity in a given usage context. We reported on a study which is composed of two user-based experiments that we set up to examine the importance of some of the listed criteria.

Given the collected results, we argue that the criteria *co-location*, *cluster*, and *hierarchy* are among the primary criteria that define how much relevant a geographic entity is — assuming that *topicality* and *spatio-temporal proximity* criteria are satisfied. The insight collected in the first experiment suggests that the proposed criterion *anchor-point proximity* is also part of these primary criteria. In turn, these results stand by the substantial difference between GR and the concept of relevance as it is commonly understood in the classic document-based IR and GIR.

Nonetheless, the two experiments presented here can only give us a first insight into GR, and the collected results are not enough to fully understand the complex system of criteria underlying GR, nor all the criteria have probably been listed. Moreover, it is essential to develop efficient and meaningful ways to communicate GR to users searching for geographic information on a mobile map (Reichenbacher, 2004; Crease and Reichenbacher, 2011). Further studies are also needed to gain a better comprehension of GR and how it can influence our understanding of personal mobility and human activity

Future research on this topic will include a validation of the results presented here through a second study, based on real-world data. We plan to develop a prototype software which will incorporate a gradual implementation of the discussed criteria. This prototype system will be used to generate GR assessments based on real-world geographic data. At first stage, the prototype system will not be sophisticated enough to deliver real-time results to a mobile user through the network. Nevertheless, given a realistic situation, a related output will be then available for further user-based experiments. This follow-up study will allow us to settle the questions left open in this paper, and to achieve an even deeper understanding of GR and the related criteria.

References

C. L. Barry. User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, 45(3):149–159, 1994. ISSN 1097-4571.

C. L. Barry and L. Schamber. Users' criteria for relevance evaluation: a cross-situational comparison. *Information Processing and Management*, 34(2-3):219–236, 1998.

Ralf Bierig and Ayşe Göker. Time location and interest: an empirical and user-centred study. In *Proceedings of the 1st international conference on Information interaction in context*, pages 79–87. ACM, October 2006 2006.

A. Bookstein. Relevance. *Journal of the American Society for Information Science*, 30(5):269–73, 1979.

B. Boyce. Beyond topicality: a two stage view of relevance and the retrieval process. *Information Processing & Management*, 18(3):105–109, 1982. ISSN 0306-4573.

P. Coppola, V. Della Mea, L. Di Gaspero, and S. Mizzaro. The concept of relevance in mobile and ubiquitous information access. In Fabio Crestani, Mark Dunlop, and Stefano Mizzaro, editors, *Mobile and Ubiquitous Information Access*, volume 2954 of *Lecture Notes in Computer Science*, pages 3–6. Springer Berlin / Heidelberg, 2004.

H. Couclelis, R.G. Golledge, N. Gale, and W. Tobler. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7(2):99–122, 1987. ISSN 0272-4944.

Paul Crease and Tumasch Reichenbacher. Adapting cartographic representations to improve the information seeking of lbs users. In *Proceedings of 25th International Cartographic Conference*, July 2011 2011.

C. da Costa Pereira, M. Dragoni, and G. Pasi. Multidimensional relevance: a new aggregation criterion. In Mohand Boughanem, Catherine Berrut, Josiane Mothe, and Chantal Soule-Dupuy, editors, *Advances in Information Retrieval*, volume 5478 of *Lecture Notes in Computer Science*, pages 264–275. Springer Berlin / Heidelberg, 2009.

S. De Sabbata. Criteria of geographic relevance. In *Proceedings of the 12th International Conference on Geographic Information Science*, September 2010 2010.

R.G. Golledge. The nature of geographic knowledge. *Annals of the Association of American Geographers*, 92(1):1–14, 2002. ISSN 1467-8306.

J. Han, M. Kamber, and A. K. H. Tung. *Geographic data mining and knowledge discovery*, chapter Spatial clustering methods in data mining: a survey, pages 188–217. Taylor and Francis, 2001.

Y. Huang, S. Shekhar, and H. Xiong. Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge and Data Engineering*, pages 1472–1485, 2004.

T. Jordan, M. Raubal, B. Gartrell, and M. Egenhofer. An affordance-based model of place in gis. In *Proceedings of 8th International Symposium on Spatial Data Handling*, pages 98–109. Citeseer, July 1998 1998.

K. Koperski and J. Han. Discovery of spatial association rules in geographic information databases. In Max Egenhofer and John Herring, editors, *Advances in Spatial Databases*, volume 951 of *Lecture Notes in Computer Science*, pages 47–66. Springer Berlin / Heidelberg, 1995.

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. ISSN 0004-5608.

S. Mizzaro. Relevance: the whole history. *Journal of the American Society for Information Science*, 48(9):810–832, 1997.

D. Mountain and A. MacFarlane. Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. *Journal of Information Science*, 33(5):515, 2007.

P.M. Pombinho, M.B. Carmo, and A.P. Afonso. Evaluation of overcluttering prevention techniques for mobile devices. In *Proceedings of 13th International Conference Information Visualisation*, pages 127–134. IEEE, July 2009 2009.

J. Raper. Geographic relevance. *Journal of Documentation*, 63(6):836–852, 2007.

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.

T. Reichenbacher. *Mobile cartography: adaptive visualisation of geographic information on mobile devices*. Verlag Dr. Hut, 2004.

T. Reichenbacher. The importance of being relevant. In *Proceedings XXII International Cartographic Conference*, July 2005 2005.

T. Reichenbacher. The concept of relevance in mobile maps. In Georg Gartner, William Cartwright, Michael P. Peterson, William Cartwright, Georg Gartner, Liqiu Meng, and Michael P. Peterson, editors, *Location Based Services and TeleCartography*, Lecture Notes in Geoinformation and Cartography, pages 231–246. Springer Berlin / Heidelberg, 2007.

T. Reichenbacher. Geographic relevance in mobile services. In *Proceedings of the 2nd International Workshop on Location and the Web*. ACM, April 2009 2009.

T. Reichenbacher, P. Crease, and S. De Sabbata. The concept of geographic relevance. In *Proceedings of the 6th International Symposium on LBS & TeleCartography* , September 2009 2009.

Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of the 28th of the international conference on Human factors in computing systems*, pages 2863–2872. ACM, April 2010 2010.

T. Saracevic. Relevance reconsidered. In *Proceedings of the 2nd Conference on Conceptions of Library and Information Science*, pages 201–218. Royal School of Librarianship, October 1996 1996.

T. Saracevic. Relevance: a review of the literature and a framework for thinking on the notion in information science. part ii: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13):1915–1933, 2007.

R. Savolainen and J. Kari. User-defined relevance criteria in web searching. *Journal of Documentation*, 62(6):685–707, 2006.

L. Schamber. Users' criteria for evaluation in multimedia information seeking and use situations. *Information Processing and Management*, 30(2), 1991.

A. Stevens and P. Coupe. Distortions in judged spatial relations. *Cognitive psychology*, 10(4):422–437, 1978.

P. Wilson. Situational relevance. *Information storage and retrieval*, 9(8):457–471, 1973.

Y. C. Xu and Z. Chen. Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology*, 57(7):961–973, 2006.

A. Zipf. Die relevanz von geoobjekten in fokuskarten. In *Proceedings of Symposium für Angewandte Geographische Informationstechnologie*, pages 567–576. Heidelberg: Wichmann, July 2003 2003.