

# Granularity as a Qualitative Concept for GIR

Dirk Ahlers  
NTNU – Norwegian University of Science and Technology  
Trondheim, Norway  
dirk.ahlers@idi.ntnu.no

## ABSTRACT

We examine the notion of granularity for qualitative thinking about geospatial data and location references. Granularity can be understood as an abstraction of level of detail or spatial resolution. Pure coordinates, which may exhibit strong overprecision for some entities, can be combined with not only hierarchical gazetteer information, but also with derived semantic data about extent of places and thus help in correct interpretations without necessarily more accuracy.

## CCS Concepts

•Information systems → Spatial-temporal systems; Information retrieval;

## Keywords

Granularity, Geographic Information Retrieval; Spatial Semantics, Spatial Data Quality

## 1. INTRODUCTION

In this position paper, we argue for an inclusion of qualitative concepts into geospatial retrieval, mainly *granularity* and related concepts such as *level of detail* or *height*. With increasing coverage and quality of gazetteers and other geospatial knowledge bases, the positional accuracy and precision of location references is continuously improving. However, in many cases, there is a gap between very high-granular references such as buildings or even indoor locations within buildings and associated mobile user positions, and of low-granularity locations such as cities or countries that provide a conceptual hierarchical foundation.

A perfect, but complex solution, would be to have full footprints or polygons available for any possible entity. However, not always are high-precision locations really necessary, but the availability of annotations for conceptual or numerical accuracy would be fully sufficient while not demanding too much computational complexity. On the other hand, related concepts of geometric granularity can help define or

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*GIR '15, November 26 - 27, 2015, Paris, France*

© 2015 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-3937-7/15/11...\$15.00

DOI: <http://dx.doi.org/10.1145/2837689.2837704>

improve additional retrieval systems. In this paper, we will develop the notion of granularity and discuss its influence in geographic retrieval.

## 2. GRANULARITY CONCEPTS

We want to make an important distinction between *locations* and *location references*. A location is an actual place physically located somewhere, while the symbolic location reference is the pointer to it or the way it gets referenced. There is a wide variety of references, both formal referencing systems as well as different levels of informal ways of toponyms or other references or descriptions. Examples are coordinates, city names, country names, zip codes, addresses, house names, parcel numbers, vernacular names, personal designations, translations, nicknames etc.

Today, there are two major connected ways to talk about location. One is coordinates, mostly in the form of GPS global coordinates and the other is the use of named places or toponyms, often in the form of gazetteers or qualified addresses useful for manual lookup or input into a navigation tool. Of course, GPS coordinates have come a long way as a unifying system to reference locations on the planet globally and are a very strong driver for mapping and analysis by providing a universal standard. However, they capture only a 'flat' aspect of location.

Thinking in coordinates can be enriched by other levels of conceptualization. Gazetteers and spatial databases are the current main way to do this. They add a hierarchy of named places as location references, which in turn are grounded by coordinates. The hierarchy usually consists of political and administrative subdivisions, but other hierarchies and facets are possible. Gazetteers are already translators between certain classes of named places and global coordinates as well as hierarchical relationships. In this way, they combine 'flat' coordinates with a hierarchical description of the world in the form of a standardized vocabulary. Another important aspect is the classification of described entities to allow reasoning about different types of entities that will also exhibit different spatial default characteristics to, for example, differentiate a city from a state, a building, or a complex campus entity by means of semantic granularity.

From the other side, that of information retrieval and analysis, semi- or unstructured texts are subjected to entity extraction to gather and disambiguate available location references to open them up to further processing. This grounding of places [5] to a commonly agreed reference, either gazetteer or coordinates is a main aspect of GIR. An interesting question is how the accuracy of the coordinate and

of the actual place relate as well as which parts and relations are implied. For system improvement, a push towards higher granularity – by increased accuracy of the coordinates – is the first idea that springs to mind. But it may not always be the only or preferred solution. Instead, we argue for a richer understanding and modelling of location.

Our definition of granularity incorporates two complementary factors, first a higher resolution of spatial references on the numerical scale of coordinates, and second, a higher resolution of the hierarchical or conceptual sub-divisions for location references. In this way, we can understand granularity also as *level of detail*. This opens an elegant way of dealing with positional inaccuracies by relating them to a conceptual hierarchy.

## 2.1 Positional and hierarchical granularity

Coordinates imply a very high granularity and accuracy, in usual coordinates (6 digits such as 63.429722; 10.393333) down to 10cm. However, the actual accuracy often is much less, either due to limitations in measuring devices, errors in transmission, or inflated accuracy when just pinpointing a place on the map. The example is the position of the city of Trondheim in Wikipedia, in which arguable the accuracy of the position shows too many significant digits in relation to the size of the city, a usual case of overprecision. Without knowing the extent or type of the entity, accuracy alone is not that useful in many applications.

Possible inaccuracies in data sources also have to be examined with this background [1]. By knowing the extent of places, uncertainty of measurements or the characterisation of coordinates as approximations can be better utilised. For example, geotags in Wikipedia can optionally be annotated with a simple zoom or size factor, which tells us something about the size (and possibly type) of a place, so that an understanding about its extent and the exactness of the coordinate is possible. Usually for larger places, no correct centroid can be assumed, but often also is not necessary. The knowledge of feature classes can give sufficient hints about the possible extent. In the case of Trondheim, knowing that it is a city and estimating size from population numbers (as is also often done in disambiguation) can get the right order of magnitude. Of course, ideally, the extent or size of a place is available. For example, footprints have been assumed as parts of gazetteers for a long time [3], but they are not always present and are often only available for entities at coarse granularity down to the country and city level or only as coarse bounding boxes [6]. However, additional sources can fill some gaps. For example, OpenStreetMap has improved a lot in this regard and now many building footprints are available worldwide.

Open issues are places within places at high granularities, for example shops within a mall, or constituent parts of a campus. Vice versa, larger entities such as campuses or industry lots may even have a richer internal structure that is important to know from the outside, for example multiple entry points into a large factory. On the other hand, we can move away from point-based locations, but also include regions, footprints, scopes, detailed shapes etc. This is inline with a general move from point data and polygons to rich 3D models. The only issue there is that they usually do not cover the whole world and thus still strong and reliably fallback solutions are important. In short, we can substitute high positional accuracy with semantic annota-

tions on the granularity to interpret coordinates according to the footprints of their associated places.

## 2.2 Using Altitude

Another way to benefit from the granularity definition is to drive the subdivision further when starting from an address or a building reference. While many addresses only specify the building, in denser areas they also routinely specify the floor of an entity. In other cases, it may not be directly described, but could be derived from available room numbers. This is a possibility for both location reference extraction and visualisation [2]. Only altitude in itself is not that useful. Only in relation to the local ground level or to other structures can we derive value. Again, conflating this with a conceptual view allows to not only define numerical height, but also define for example the floor number in a multi-storey building. Yet, in many large-scale structures or regions, local elevation does not play such a major role for many applications and it would rather be of interest for smaller structure. Accordingly, for example in GeoNames, less than 1% of populated places have elevation data.

Recent work has begun to show some departure from purely flat coordinates, for example for points of interest (POIs) while of course mapping and visualisation solutions such as Google Earth have been able to view topological terrain height as well as 3D building models for some time. Seeing the third dimension also being considered for referencing is a valuable step.

## 3. CONCLUSION AND FUTURE WORK

It is clear that we can think about location in multiple layers and concepts. The question is then, how can they be connected? Granularity is a handy concept to think about location references. We have started to use the notion of granularity in the development of a conceptual framework that makes it easier to deal with location at different scales.

We aim to refine and apply this concept in future work. In some cases, granularity annotations alone may not be enough, for example when talking about complex spaces, such as dispersed pieces of the campus of a university and more complex polygons are needed [4]. Access limitations for high positional accuracy can be mitigated with annotations or derivations of the granularity levels to interpret coordinates according to their associated places.

## 4. REFERENCES

- [1] D. Ahlers. Assessment of the Accuracy of GeoNames Gazetteer Data. *GIR '13*. ACM, 2013.
- [2] D. Ahlers. Search and navigation in complex overlapping urban spaces. In *Making Places Workshop @ NordiCHI2014*, 2014.
- [3] L. L. Hill. Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. ECDL, 2000.
- [4] B. Martins, F. J. López-Pellicer, and D. Ahlers. Expanding the Utility of Geospatial Knowledge Bases by Linking Concepts to WikiText and Polygonal Boundaries. In *GIR '15*, 2015.
- [5] S. E. Overell and S. M. Rürger. Identifying and grounding descriptions of places. In *GIR 2006*, 2006.
- [6] C. Schlieder, T. Vögele, and U. Visser. Qualitative spatial representation for information retrieval by gazetteers. In *COSIT 2001*. Springer, 2001.