Linking IOT and GIS using GeoHashes to Create a Competitive Advantage

Laurie Butgereit

School of ICT
Nelson Mandela University
Port Elizabeth, South Africa
laurie.butgereit@mandela.ac.za

***Abstract*—The Internet of Things enables physical objects to be first class citizens of the digital world. The various Internet of Things technologies give physical objects the ability to provide information and react to external instructions. Geographical Information Systems provide analytics to spatial data. The use of GeoHashes creates a partnership between the Internet of Things and Geographical Information Systems which can provide a competitive advantage to businesses. This paper describes a Design Science Research project to create an artifact which enables geographical intelligence gathering and analytics in order to better compete in the business environment.**

Keywords*—*Internet of Things, Geographical Information Systems, GeoHashes

I. Introduction

"If you know the enemy and know yourself, you need not fear the result of a hundred battles" [1] So said Sun Tzu in his epic work *The Art of War*. According to a Forbes magazine article, these words are as true now in a business environment as they are in a military environment [2]. Put in a commercial context, if a business understands its challengers and if a business understands its own operations, then the business will be successful in a competitive market place.

This paper looks at how linking Internet of Things technology together with Geographical Information Systems through the use of GeoHashes can provide competitive advantage to businesses.

The term *Internet of Things* describes a situation where physical objects interact seamlessly with the digital world (This is more clearly described in Section II.A). The term *Geographical Information Systems* embodies many traditional Information Systems techniques which are specifically tailored for handling spatial information (This is more clearly described in Section II.B). The term *GeoHashes* describes a technique for storing geographical information (latitudes and longitudes) in a string in which locations which are nearby to each other also sort close to each other (This is more clearly described in Section II.C).

This paper provides a brief literature review of the three individual topics (Internet of Things, Geographical Information Systems and GeoHashes) in Section II. A description of the research environment is provided in Section III. Design Science Research was used for this project and an argument of how Design Science Research and Agile methodologies are compatible is provided in Section IV.

This research created an Instantiation as defined by Design Science Research and this Instantiation is described in Section V. The results of the research and the conclusions can be found in Sections VI and VIII respectively. The ethics which guided this research is described in Section VII

As can be seen by the description of the research enviornment in Section III, this research was conducted at a business entity. As such, some aspects of this research are considered private intellectual property of the company under research and can not be described in full in this paper. This paper only deals with the aspects of the research which the company deemed to be shared with the greater public.

II. Literature Review

This section provides a brief overview of the existing literature in the areas of Internet of Things, Geographical Information Systems, and GeoHashes. Literature and definitions on the Internet of Things can be found in Subsection A. Literature and definitions on Geographical Information Systems can be found in Subsection B. Literature and definitions on GeoHashes can be found in Subsection C.

A. Internet of Things

Haller *et al* define the *Internet of Things* as "a world where physical objects are seamlessly integrated into the information network, and where the physical objects can become active participants in business processes" [3].

A 2015 report by McKinsey states that by 2025 the value of the Internet of Things could generate over ten trillion US Dollars per annum. Before this economic value can be obtained, however, a number of problems need to be resolved including interoperability [4].

B. Geographical Information Systems

Many attempts have been made to define the expression *Geographical Information Systems*. As far back as in 1991, Mcguire attempted to collate a wide variety of definitions into one unified definition [5]. He explains that Geographical Information Systems live at the intersection of four disciplines: 1) Computer cartography 2) Remote sensing 3) Database Management Systems and 4) Computer aided design. From the literature which Mcguire cites, he determined that there are three views of Geographical Information Systems 1) the map view 2) the database view and 3) the spatial analytical view.

The map view focuses on the cartographical aspects of Geographical Information Systems. The database view focuses on a well designed and implemented database. The third view emphasises the importance of spatial analysis.

C. GeoHashes

GeoHashes were invented by Gustav Niemeyer in 2008 [6]. Niemeyer himself does not appear to publish peer-reviewed research but in a 2008 forum post at geocaching.com, Niemeyer announced his initial work with GeoHashes [7]. In this forum post, Niemeyer explains briefly that a pair of latitude/longitude values can be encoded into a short string in such a way that nearby places have similar GeoHash prefixes. That is to say that the strings representing the locations of nearby places will sort near to each other alphabetically.

There appears to be no official document explaining how Niemeyer released the GeoHash algorithm into the public domain although many non-peer-reviewed sites explain that it was concurrent with the creation of his website geohash.org. Regardless, there are now numerous public domain implementations of the GeoHash algorithm.

III. Project Environment

This project was done at a public company listed on the Johannesburg stock exchange. The company maintains a network of tens of thousands of low cost Android devices which are provided to merchants around the country. The merchants range in size from small one-person spaza shops to large chain department stores and petrol stations.

The Android devices are restricted in use and run an application written by the company which enables the sale of virtual digital products such as cell phone airtime, cell phone data, wifi connectivity, and prepaid electricity. In addition, products such as metered electricity, topup airtime, topup data, bus tickets, event tickets, bill payments, etc, are facilititated on the device. The Android devices connect to a proprietry network maintained by the company. The network maintains onward connections to the actual service providers such as the cell phone companies, bus companies, etc.

Previous work by the company included the embedding of device information inside of the commercial transactions between the Android devices and the company network. These embedded messages allowed the company to take part in the Internet-of-Things as defined by Haller *et al* [3].

These embedded messages included the longitude and latitude of the device, identifiers of nearby cell towers, SIM card serial number, device serial number, device model, identifier of any nearby wifi hotspots, etc. This information was embedded in a subset of commercial transactional messages and was sent frequently and seemlessly to the company network. When the commercial transaction was processed, the embedded information was separated from the commercial information and the two different sets of information continued on two separate data paths.

Previous work by the company also included the display of these devices on various types of maps. This included real-time display of sales on dynamic maps, static displays on static maps, geographical clustering, and geographical search criteria. All of these fall into the definition of Geographical Information Systems as defined by [5].

Over time, however, the number of devices increased. At its initial stages, when a geographical search was made, all devices were extracted from the database and simple distance calculations were done to determine which devices were inside the search criteria. In addition, over time the type of search queries changes from internally facing such as “find all devices in Johannesburg” or “find all devices belonging to merchant X” to more externally facing such as “find all devices near branches of competitor X” or “find all devices near football stadia”.

It was the combination of more devices being tracked and the wider variety of search queries which was the impetus behind the research defined in this paper.

IV. Research Methodology

Prior to the research described in this paper, the company under research had embarked on a transformation to Agile software development methodologies. These Agile methodologies aligned nicely with Design Science Research which was used for this particular research project.

Design Science Research is an iterative methodology consisting of 5 steps: 1) awareness 2) suggestion 3) development 4) evaluation 5) conclusion [8], [9]. At the evaluation step, if the results are not yet satisfactory or complete, the methodology loops back to a previous step. This aligns well with Agile Principles 1 through 4 which state [10]

1. Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2. [We] welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
3. [We] deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4. Business people and developers must work together daily throughout the project.

By adhereing to the first four Agile Principles, the research also adhered to the 5 step iteration cycle of Design Science Research.

Design Science Research also supports three interleaving cycles:

1. Relevance Cycle – This cycle ensures that the research attempts to solve a real world problem
2. Rigor Cycle – This cycle ensures that the research is grounded in scientific and technical excellence
3. Design Cycle – This cycle is the 5 step iteration cycle previous described

These cycles can be seen in Illustration 1

These three cycles also align with the principles of the Agile Manifesto. The daily meetings described in Agile Principle #4 along with Agile Principle #6 (“The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.”) ensure that the Relevance cycle is adhered to. Agile Principle #9 (“Continuous attention to technical excellence and good design enhances agility”) ensures that the Rigor cycle is adhered to. And Agile Principles 1-4 as already itemised ensure that the Design cycle is adhered to.

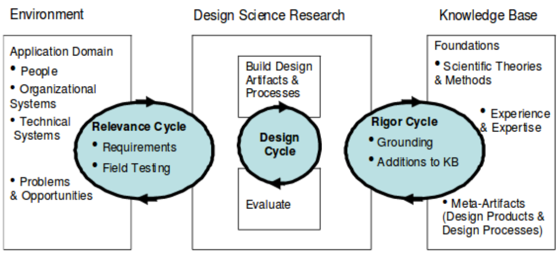

Illustration 1: Three Cycles of Design Science Research (Hevner)

Design Science Research supports a number of different types of research outputs including (but not limited to):

1. Constructs – an underlying vocabulary of a new research area
2. Methods – algorithms and techniques for solving specific problems
3. Models – designs for solving larger problems
4. Instantiations – implementations of models and methods

The research described in this paper created an Instantiation on how the Internet of Things and Geograpahical Information Systems can be joined together using GeoHashes in order to provide business advantages over competitors.

V. Instantiation

This section describes the Design Science Research Instantiation which was created by this project. This section is divided into subsections which describe the overall architecture of the solution (Section V.A) and the low level technicalities of the solution (Section V.B)

A. *Overall Architecture*

An overview of the architecture can be seen in Illustration 2. When the devices send commercial messages to the company under research, the spatial data which is embedded in the messages is extracted. Commercial information is sent along normal business lines of communication and is not within the scope of the paper. The spatial data, however, continues along a separate data path. The latitude and longitude of the individual devices is stored and the GeoHash of each device is also encoded as a string and stored.

There is a separate Graphical User Interface (GUI) access to the spatial data. In addition, however, the GUI allowed the user to search the broader Internet for spatial information. For example, it might be important to know the location of bank branches. Banks freely publish this information on their websites along with the addresses of their branches. This information could be scraped from websites. The actual physical addresses of the the branches could be stored, geocoded and their GeoHashes could be calculated and stored.

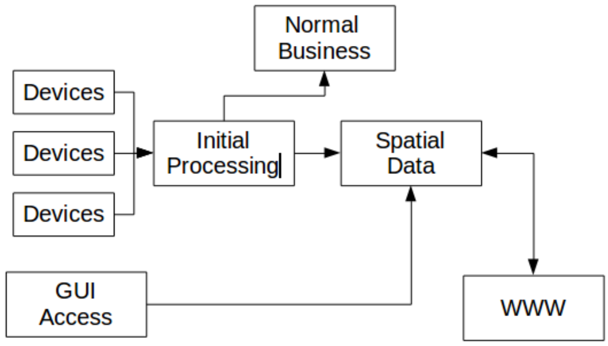

Illustration2: Architectural Overview

This combination of company private information (location of individual devices) combined with spatial information about external entities (for example, location of all the bank branches) provides additional intelligence. This combination now makes it possible to query for all devices which are within N kilometers of a bank branch. Prior to the use of GeoHashes, this search would require mathematical calculations for each comparison. The use of GeoHashes ensure that this calculation is no longer necessary and mere string sorts could be used.

The output of this research could either be viewed on the screen, printed on paper maps, or directly sent to the company outbound call center to initiate verbal human-to-human communicate with the merchant owning the device which had been identified.

B. *Technicalities*

This instantiation was developed in Java using the SpringBoot framework. Hibernate provided an object oriented view of an SQL database. PrimeFaces provided the actual web interface.

Visual mapping along with geocoding was provided by paid subscription to the various Google mapping libraries. This included Java Script dynamic maps, static maps, both geocoding and geolocating, locations of roads, and generation of driving instructions.

A separate public domain GeoHash implementation written by David Moten [11] was used for the encoding and decoding of GeoHashes.

VI. Results

A number of different types of business research problems were the basis of the testing of this Instantiation. Although the exact business research is company private, two similar examples are given in this section. The first example deals with devices located near well known landmarks. The second example deals with devices located near entities deemed to be competitors.

The first test of the Instantiation was whether or not devices located within a specific distance of various types of landmarks could be located. For example a hypothetical task could be to locate all devices within N kilometers to a football stadium. The location of the football stadia could easily be found on the public internet. The stadia addresses could be geocoded into a latitude/longitude pair using the various Google paid libraries. A GeoHash could then be calculated for each latitude/longitude pair.

All of the GeoHashes of devices and stadia could be sorted together. Depending on the required distance from a stadium, devices whose GeoHash sorted near the GeoHash of a stadium could be targeted by either telephone calls from the company call center or visits by sales representatives.

The second test of the Instantiation was whether or not devices could be located which were nearby competing entities. If a list of competing entities could be found on the public internet (See section VII regarding the ethics of such searches), then these entities could also be geocoded and a GeoHash of their location also calculated. The same technique could be used with the location of well known landmarks.

VII. Ethics

This section describes the ethics governing this research

No personal information was required of any private individuals during the scope of the research. During the development of the Instantiation under question, some public domain utilities were used (such as the GeoHash algorithm) and some private utilities were used under license (such as the various Google utilities). There were no ethical issues regarding the actual development of the Instantiation.

During testing and evaluating the instantiation, however, various websites were scraped for their information. The term *scraped* or *web scraping* refers to the extraction or harvesting of information from pages on the Internet and the subsequent saving of this information in a database for subsequent analysis [12].

The use of web scraping, however, is controversial and may raise legal questions with regard to copyright [13]. However, Zhao cited above maintains that “An ethical web scraping tool will avoid this [copyright] issue by maintaining a reasonable requesting frequency” [12]. In view of this, the websites which were scraped were only scraped once, the physical addresses were obtained, geocoded, and geohashed. In addition, only publicly available data which was visible on the Internet without any password protection was ever scraped for this research.

VIII. Conclusion

Sun Tzu said “If you know the enemy and know yourself, you need not fear the result of a hundred battles.”

The first portion of Tzu’s quotation is “If you know the enemy...” In the research described in this paper, this phrase is the equivalent of finding spatial information about competitors. This spatial information could be merely physical addresses which are freely provided on the open internet (such as bank branches). This spatial information could be geocoded and GeoHashes could be calculated and stored.

The second portion of Tzu’s quotation is “...and know yourself.” In the research described in this paper, this phrase is the equivalent of ensuring that all the spatial information about internal company assets is stored, geocoded, and GeoHashed.

Competitive business can be conceived to be a battle between different companies. Assuming that both the first two portions of Tzu’s quotation has been satisfied, then “...you need not fear the result of a hundred battles.”

References

[1] S. Tzu, "The Art of War (Translated from the Chinese by Lionel Giles)", Accessed Apr. 26, 2020 [Online] http://classics.mit.edu/Tzu/artwar.html.

[2] Jackson E., "Sun Tzu's 31 Best Pieces of Leadership Advice," *Forbes,* May 23, 2014.

[3] S. Haller, S. Kamouoskos and C. Schroth, "The Internet of Things in an enterprise context," *Future Internet Symposium,* pp. 14-28, 2008.

[4] J. Manyika *et al*, "Unlocking the Potential of the Internet of Things," *McKinsey Global Institute,* 2015.

[5] D. J. Maguire, "An overview and definition of GIS," *Geographical Information Systems: Principles and Applications,* vol. 1, pp. 9-20, 1991.

[6] A. Fox *et al*, "Spatio-temporal indexing in non-relational distributed databases," in *2013 IEEE International Conference on Big Data,* 2013, .

[7] G. Niemeyer, *Post on forums of Geocaching.com,* Feb. 6, 2008, Accessed on Apr. 26, 2020 [Online] Available: https://forums.geocaching.com/GC/index.php?/topic/186412-geohashorg/.

[8] A. R. Hevner *et al*, "Design science in information systems research," *Management Information Systems Quarterly,* vol. 28, *(1),* pp. 75-106, 2004.

[9] A. R. Hevner and S. Chatterjee, *Design Research in Information Systems: Theory and Practice.* 2010.

[10] Agile Alliance, *Agile Alliance Website*, 2020, Accessed on Apr. 26, 2020 [Online] Available: https://www.agilealliance.org/.

[11] D. Moten, *GeoHash algorithm GitHub page,* Aug. 3, 2017, Accessed on Apr 26, 2020 [Online] Available: https://github.com/davidmoten/geo.

[12] B. Zhao, "Web scraping," *Encyclopedia of Big Data,* pp. 1-3, 2017.

[13] S. O'Reilly, "Nominative fair use and Internet aggregators: Copyright and trademark challenges posed by bots, web crawlers and screen-scraping technologies," *Loy.Consumer L.Rev.,* vol. 19, pp. 273, 2006.