Using GeoHashes to Combine IOT and GIS to Provide Business Intelligence to the Informal Sector in South Africa

Laurie Butgereit

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

Abstract

South Africa has one of the highest GINI coeefficient indicating a high degree of inequality in the country. There is also extreme unemployment with the expanded unemployment rate being 38.3% and in some subsections of the economy as high as 68.3%. Despite this, the Informal Sector (non-agricultural) employs over three million people. Many corporates offer products to the formal sector, the informal sector or both. The commercial margins are often very slim in the informal sector. This paper looks at the use of Internet of Things, Geographical Information Systems, and GeoHashes to provide business intelligence to merchants in the Informal Sector thereby helping them improve their competitive advantage.

1 Introduction

According to a recent World Bank report, South Africa is one of the most economically unequal countries in the world with a consumption expenditure GINI coefficient of 0.63 . The richest 10% held 71% of the net wealth in 2015 and the bottom 60% held 7% of the net wealth [1]. The Quarterly Labour Report for Q2 2019 published by Statistics South Africa (the official purveyor of statistical information) put the unemployment rate at 29%. The expanded unemployment rate which measures people who have stopped actively looking for work is as high as 38.3%. In some subsections of the economy, for example young people aged 15-24, the expanded unemployment rate is as high as 68.3% [2].

There is a vibrant Informal Sector in the economy, however. The Q2 2019 Labour Report also shows over 3 million people employed in the informal market (non-agricultural).

Many corporates in South Africa offer products and services geared to the formal sector, the informal sector or both sectors. The commercial margins in the informal sector can be very small.

For example, some of the virtual digital products described in this paper (Section 4) have a full retail value of less than the equivalent of fifty US cents. Mechanisms need to be in place so that merchants can sell such products, make a profit, and continue to stay in business. In addition, the corporates involved need to also make a profit for these services to continue.

This paper specifically looks at using GeoHashes to merge the Internet-of-Things with various Geographical-Information-Systems in order to give business intelligence to vendors and merchants working in the Informal Sector thereby improving their competitive advantage.

Definitions are provided in Section 2. Section 3 describes the research methodology used in this project. Section 4 describes the actual environment in which the research took place. Section 5 describes the actual artifact which was created and its underlying architecture. The results and conclusions are found in Sections Error: Reference source not found and 7 respectively.

It is noted that some of the information about this research is considered company private intellectual property and will not be disclosed in this paper.

2 Definitions and Literature

The expression *Internet of Things* has been defined 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, McKinsey report 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].

There have been many attempts to define the expression *Geographical Information Systems.* One of the most extensive attempts was by McQuire in 1991 in which he attempted to create one unified definition [5]. He explained that Geographical Information Systems live at the intersection of 1) Computer Cartography 2) Remote sensing 3) Database management and 4) computer aided design. McQuire determined that there are three views of Geographical information Systems 1) map views (which focus on the cartographical aspects) 2) database views (which focus on well designed and implemented databases) and 3) spatial analytical views (which emphasize the importance of spatial analysis).

Gustav Niemeyer invented GeoHashes in 2008 [6]. Although Niemeyer himself does not appear to publish peer-reviewed research, in a 2008 forum post at geocaching.com, Niemeyer announced his initial work with GeoHashes [7]. In this forum post, Niemeyer explains how a pair of latitude/longitude values can be encoded into a short string in such a way that nearby places have alphabetically similar GeoHash prefixes. That would mean that GeoHash strings which sort close to each other alphabetically are also geographically close to each other. Niemeyer released the GeoHash algorithm into the public domain concurrently with the creation of his website geohash.org. There are numerous open source implementations of the GeoHash algorithm.

Business Intelligence can be defined as a path to lower costs, improved service delivery, and better decision making. Business Intelligence provides business owners and managers tools and data for making informed decisions [8]. Although the term has been in existence for many years, in more recent times the term has become related to data science. The input to Business Intelligence projects is data. The output from such projects is information [9].

According to Google Books Ngram viewer, the terms *informal sector* or *informal economy* only started to become in widespread use around 1970. In his 1973 paper, Hart described *informal income* opportunities in Ghana. He explained how various economic factors "...led to a high degree of informality in the income-generating activities" of the residents [10]. In a later article, Hart explained that the term arose in response to the growth of self-employment and casual labour in Third World cities [11].

This paper describes a project in providing business intelligence to merchants in the informal sector by using GeoHashes to link information from the Internet of Things and Geographical Information Systems.

3 Research Methodology

Design Science Research was used for this project because it aligned nicely with various Agile methodologies which were being used in the company under research.

Design Science Research consists of three interwoven cycles as can be seen in Illustration 1 [12], [13]. In the Relevance Cycle, the researcher interacts with the application domain. This aligns well with the daily meetings mandated by Agile Principle #4 ("Business people and developers must work together daily throughout the project") and the face-to-face conversations mandated by Agile Principle #6 ("The most efficient and effective method of conveying information to and within a development team is face-to-face conversation") [14]. In the Rigor Cycle, the researcher interacts the the scientific theories and knowledge base. This aligns well with Agile Principle #9 ("Continuous attention to technical excellence and good design enhances agility") which mandates attention to technical excellence. In the Design Cycle, there is a five step iterative loop of 1) Awareness 2) Suggestions 3) Development 4) Evaluation and 5) Conclusion.   This iterative loop aligns well with the short development cycles mandated by Agile Principles #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."), and #3 ("[We] deliver working software frequently, from a couple of weeks to a couple of months, with preference to the shorter timescale").

There are a number of possible outputs of Design Science Research including constructs, methods, models, and instantiations. Constructs provide the language in which problems and solutions can be defined and communicated within a research domain. Methods define algorithms, "best practices" and solution processing. Models represent real world solutions. Instantiations are real world implementations of methods and/or models.

This research described in this paper presents an Instantiation of a Model where GeoHashes can be used to merge Internet of Things and Geographical Information Systems. This Instantiation was implemented in the Informal Sector of South Africa and provided business intelligence to merchants in the Informal Sector.
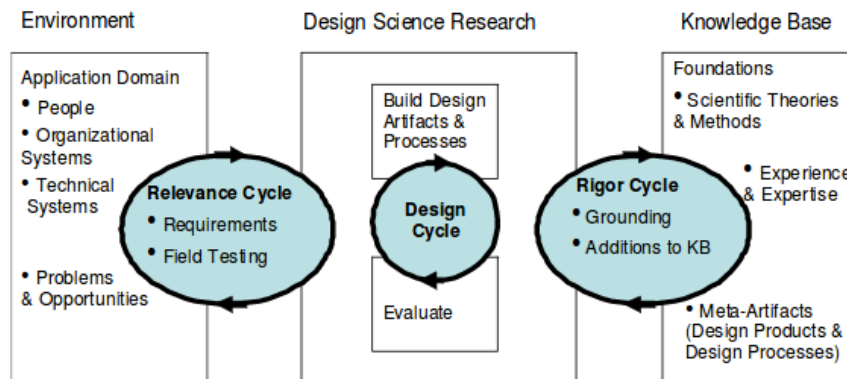
Illustration 1: Three cycles of Design Science Research (Hevner)

4 Environment

This research was done at a public company listed on the Johannesburg Stock Exchange. The company is one of the largest suppliers of secure tokens in South Africa which can be exchanged for various digital products such as cell phone airtime, cell phone data, and electricity. The company also offers other non-tangible products such as bus ticketing, event ticketing, bill payment facilitation, and traffic infringement payment. It operates in both the formal sector and the informal sector of South Africa. This paper will now only deal with the Informal Sector.

The sale of the various digital products is facilitated through a network of inexpensive Android devices which are provided to merchants at no up front cost. The devices run a proprietary app developed by the company under research. This app connects to a proprietary network operated by the company and that network connects onward to the various service providers such as cell phone companies, bus companies, and electricity providers.

The commercials offered by the company make it feasible for merchants in the Informal Sector to receive a device free of charge and start selling various digital products immediately. The various terms of the commercial agreements are not part of the scope of this research.

The Informal Sector in South Africa is typically either in densely populated townships (as in Illustration 2) or in sparsely populated rural areas.

Although Illustration 2 shows an area of Alexandra Township which was actually traversed by the Google Street View car and is well mapped, many more of the townships in South Africa are not mapped well by Google and the Google geocoding API merely returns the string "Unamed Road" in the formatted address field.

This research is about providing business intelligence to merchants in both of these informal areas. This business intelligence helps enable these merchants to better compete in the open market.

South Africa has eleven official languages and even more unofficial ones. And although there is good work being done in automated translation via machine learning [15], this research culminates in messages through a corporate call center. The business intelligence is provided to the merchants through human-to-human conversations in a handful of human languages.

Illustration 2: Alexandra Township (Google map satellite view)

5 Artifact

None of the topics in the title of this paper are particularly new. The Internet of Things is not new. Geographical Information Systems are not new. GeoHashes are not new. Business Intelligence is not new. The Informal Sector is not new

This research however combines these features in an innovative manner to create an Instantiation which provides business intelligence to merchants in the Informal sector (both rural and urban) in a cost effective manner by harnessing the power of the Internet of Things and Geographical Information Systems joined through GeoHashes.

At the heart of the Artifact created in this research is the embedded information from the tens of thousands of Android devices which are distributed throughout the country. Periodically, the devices embed information such as the longitude and latitude of the device, cell tower identifiers of nearby cell towers, device serial numbers, model numbers, etc, into the body of the commercial transaction which they send to corporate servers. When the commercial transaction is received, the device information is separated from the commercial information and the two sets of information continue down two separate data paths. The path of the commercial data is not part of this research and will not be discussed. An overview of this dataflow can be seen in Illustration 3.

When the location data is extracted from the commercial message, the GeoHash is calculated and stored.

In addition, other sources of interesting features are extracted from various web pages which can be found. This includes features such as locations of bus stops, locations of taxi ranks, locations of competitors, locations of schools, etc. This is an ongoing process and the GeoHashes of all of those interesting features are also calculated and stored. The Google Places API (a paid API service) is also queried for places of interest nearby devices such as banks and deposit-accepting ATMs (Automated Teller Machines).

At times throughout the business cycle, it is beneficial to contact merchants who are nearby certain features. For example, when a new bus carrier is added to the various digital products which can be sold, then it is beneficial that all the merchants within a certain distance of the bus stops of this new bus carrier need to know this in order for them to market their services. By doing string searches for GeoHashes with similar prefixes, all the devices which are close by the appropriate bus stops can

be found. Another example would be if a new electricity provider in a specific township or area became available, then merchants in that area need to know that they can now sell digital tokens for that electricity provider.

Merchants in the Informal Sector in South Africa often employ a number of people. It was not advantageous to just send messages to the actual commercial devices even if the human language barrier could be overcome. Once the list of devices which need to be contacted is determined, the list is then sent to the telephony system of the corporate outbound call center. As mentioned, South Africa has eleven official language and the call center employs people with a wide knowledge of local languages. In addition, although the merchant himself may be fully literate, the employees may not be fully literate and interaction with a human call center operator is more successful and textual messages.
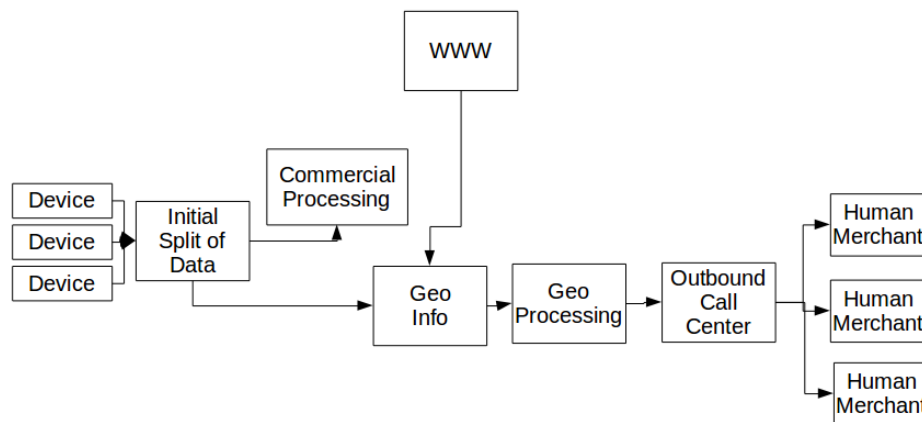

Illustration 3: Overview of data flow

6 Results

In this project, the Internet of Things provided the physical location as longitude and latitude of each of the tens of thousands of devices operated by the company under research in the Informal areas. These longitudes and latitudes were collected multiple times per day and their GeoHashes were calculated. In addition, Geographical Information Systems included the searching for interesting features on the internet or through use of the Google Places API. These interesting features included locations of football stadia, locations of banks, bus stops, taxi ranks, competitors, etc. The physical addresses of the interesting features were geocoded and the GeoHashes calculated.

Depending on business requirements, it was now possible to make generalized spatial queries. Although the exact queries are not shared in this paper, similar queries will be. If, for example, a list of competitors and their addresses could be found, then merchants or devices which were nearby these competitors could be identified by sorting the GeoHashes and looking for devices whose GeoHash sorted nearby the GeoHash of a competitor. The information about these merchants such as their names and telephone numbers could be send to the output call center. Human call center operators would then phone the merchants and have human-to-human conversations with them about how best to compete against these entities.

7 Conclusion and Going Forward

South Africa is one of the most economically unequal countries in the world. Employment publications by Statistics South Africa divide the national economy into four areas 1) formal sector (non-agricultural) 2) informal sector (non-agricultural) 3) agriculture 4) and private households. The informal sector is vibrant and employs millions of people.

The company under research described in this paper deals in both the formal sector and the informal sector. This paper only deals with interactions with the informal sector. The company provides low cost sales devices which periodically during the day embed environmental information (including longitude and latitude) inside their commercial transactions. This allows the company to keep track of the exact geolocation of tens of thousands of devices which are spread out through South Africa. A GeoHash is calculated for each of these devices from their individual longitude and latitude pairs.

In addition, the wider internet is searched for interesting geographical features throughout the country which may influence merchants operating in the informal market and their surrounding potential customers. These features include such things as sports arenas, banks, bus stops, taxi ranks and hospitals. The GeoHash of all of these features is also calculated.

The two sets of GeoHashes can be sorted together so that merchants which are nearby specific types of features can be identified.

This provides a way to identify merchants which may benefit from business intelligence based on these geographical features. Typically, business intelligence is a data driven process and is not necessarily available to informal merchants. By having the geographical information systems and business intelligence systems executing on corporate servers and communicating the output through a human call center, merchants in informal areas could benefit. For example, merchants within a short distance to football stadia may benefit from discounted surplus tickets an hour before the match begins. Merchants within a specific township may benefit from marketing information about a new electricity service provider which has started accepting secure tokens in that area.

Going forward, the company under research is adding to its database of interesting geographical features which influence economic activity in the informal market. This is an ongoing process with the sales and marketing teams making recommendations and requests for new feature sets.

Although there is no immediate plans to implement automated translations into local languages, it is not without its possibilities.

By providing relevant business intelligence to informal merchants depending on geographical conditions, these informal merchants can continue to contribute to the millions of people who are employed in the informal market in South Africa.

References

[1] (Oct 10, 2019). *The World Bank in South Africa*. Available:
https://www.worldbank.org/en/country/southafrica/overview#1.
[2] Anonymous "Quarter labour force survey, quarter 2, 2019," Statistics SA, Pretoria, South Africa, Tech. Rep. P0211, July 30, 2019. 2019.
[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] (Feb 26, 2008). *Post on forums of Geocaching.com*. Available: https://forums.geocaching.com/GC/index.php?/topic/186412-geohashorg/.

[8] É. Foley and M. G. Guillemette, "What is business intelligence?" *International Journal of Business Intelligence Research (IJBIR)*, vol. 1, *(4)*, pp. 1-28, 2010.

[9] D. Larson and V. Chang, "A review and future direction of agile, business intelligence, analytics and data science," *Int. J. Inf. Manage.*, vol. 36, *(5)*, pp. 700-710, 2016.

[10] K. Hart, "Informal income opportunities and urban employment in Ghana," *The Journal of Modern African Studies*, vol. 11, *(1)*, pp. 61-89, 1973.

[11] K. Hart, "The informal economy," *Cambridge Anthropology*, pp. 54-58, 1985.

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

[13] A. R. Hevner and S. Chatterjee, *Design Research in Information Systems: Theory and Practice.* 2010*22*.

[14] (2017). *Agile Alliance Website*. Available: https://www.agilealliance.org/.

[15] J. Z. Abbott and L. Martinus, "Towards neural machine translation for African languages," *ArXiv Preprint arXiv:1811.05467*, 2018.