International Journal of
Geo-Information

Article

Employing Search Engine Optimization (SEO) Techniques for Improving the Discovery of Geospatial Resources on the Web

Samy Katumba [1,2] and Serena Coetzee [2,*]

[1] Gauteng City-Region Observatory, a Partnership of the University of Johannesburg, the University of the Witwatersrand (Johannesburg), the Gauteng Provincial Government and Organized local Government in Gauteng, Private Bag 3, Wits, Johannesburg 2050, South Africa; samy.katumba@gcro.ac.za

[2] Centre for Geoinformation Science, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Lynnwood Road, Pretoria 0083, South Africa

* Correspondence: serena.coetzee@up.ac.za; Tel.: +27-12-420-3823

Received: 20 July 2017; Accepted: 29 August 2017; Published: 7 September 2017

Abstract: With the increasing use of geographical information and technology in a variety of knowledge domains and disciplines, the need to discover and access suitable geospatial data is imperative. Most spatial data infrastructures (SDI) provide geoportals as entry points to the SDI through which geospatial data are disseminated and shared. Geoportals are often known in geoinformation communities only, and they present technological challenges for indexing by web search engines. To overcome these challenges, we identified and categorized search terms typically employed by users when looking for geospatial resources on the Web. Guided by these terms, we published metadata about geospatial sources "directly" on the Web and performed empirical tests with search engine optimization (SEO) techniques. Two sets of HTML pages were prepared and registered with Google and Bing respectively. The metadata in one set was marked up with Dublin Core, the other with Schema.org. Analysis of the results shows that Google was more effective than Bing in retrieving the pages. Pages marked up with Schema.org were more effectively retrieved than those marked up with Dublin Core. The statistical results were significant in most of the tests performed. This research confirms that pages marked up with Schema.org and Dublin Core are a novel alternative for improving the visibility and facilitating the discovery of geospatial resources on the Web.

Keywords: geospatial resource discovery; geospatial metadata; geospatial data; spatial data infrastructure (SDI); search engine optimisation (SEO); schema.org; Dublin Core

1. Introduction

Geographic information has become ubiquitous. Every day, people from a variety of knowledge domains and disciplines use geographic information on desktop computers, mobile devices or over the Internet (web) to answer questions related to location: "Where?". Answers provide solutions in transportation planning, logistics, business, environmental monitoring, natural resource management or mining, to name a few. Providing the answers depends on suitable geospatial data, ready to be processed, i.e., geospatial data needs to be discoverable and available.

Initiatives geared towards making geospatial data available have resulted in the implementation of spatial data infrastructures (SDIs), a blend of legislation, institutional arrangements (policies), people and technologies to assure the availability, access and sharing of geospatial data [1]. An SDI enables geospatial data producers to share their data in a geoportal for discovery, access and use. Geoportals are web platforms that allow searching for geospatial data and associated metadata.

However, business, legal and technological barriers make the geospatial data in a geoportal "invisible" to the general public and to general-purpose web search engines [2]. To discover geospatial data, interested parties have to know where and how to access the geoportal or catalogue web service providing information about data. People in geographic information communities may know this, but what about those outside these communities who want to use geospatial data in their domains of expertise? Furthermore, current geoportals are mainly built on the Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW) [3]. The CSW provides an HTTP binding, designed to enable "the discovery and retrieval of spatial data and services metadata" [4]; it was not designed to be crawled by web search engines [5] and is therefore part of the "Deep Web", i.e., online content inaccessible to Web crawlers.

In response to the technological barriers described above, focused web search engines (crawlers) have been developed with the sole mission of retrieving geospatial resources (services and data) hidden behind geoportals. Such efforts are focused on enhancing the web crawler's capabilities to discover and understand geospatial resources in a geoportal [6,7]. They do not exploit the ability of such resources to make themselves visible to web search engines. In many cases, these focused crawlers use web pages retrieved by Bing, Google or Yahoo as seeds for further search refinement [2].

According to the Pew Internet Project, in America (USA), close to 92% of all activities on the Internet involve the use of search engines [8]. An estimated 3 billion searches per day were conducted on Google in 2009; by 2016 the figure has gone up to 5.5 billion [9]. For most people, Google has become not only the first port of entry for information seeking, but the only port of call [10]. An alternative approach to geospatial resource discovery is therefore to publish geospatial metadata contained in CSW catalogues "directly" on the Web. This has been done for Dublin Core metadata in Resource Description Framework (RDF) format, following the linked data principles [5]. However, the retrieval effectiveness of different vocabularies with different search engines has not yet been evaluated. Search engine optimization (SEO) techniques, which are applied in mainstream online information retrieval, have not yet been explored for the discovery of geospatial data.

The research presented in this article aims to address this gap with empirical research about the discoverability of web pages with geospatial metadata content by general-purpose web search engines. We used SEO techniques, which are widely used as a strategy for enhancing the visibility of web pages in the results of search engines [11–14]. Improving the visibility of web pages with geospatial metadata can have a significant impact on the discoverability of geospatial resources. In our research, we identified and categorized search terms typically employed by users when searching for geospatial data on the web using general-purpose web search engines. Guided by these terms, we published pages with information about geospatial data (resources) on the Web and compared the retrieval effectiveness of two metadata vocabularies (Dublin Core and Schema.org) with two web search engines (Google and Bing).

The remainder of the article is structured as follows: Section 2 provides background and further justification for the research; Section 3 describes the methods for the experiments; Section 4 presents the results, which are discussed in Section 5. Concluding remarks are also in Section 5.

2. Background and Related Work

2.1. Search Engine Optimization (SEO) Techniques for Visibility of Web Resources

Since its inception, the World Wide Web has seen a growing number of web search engine users and the proliferation of websites and web pages. Web page visibility has always been at the forefront in the design and implementation of web search engines. When tuning web pages to improve their visibility on the web, various techniques are adopted depending on the type and nature of the information being shared, as well as the kind of business or commercial activity the web page owners are involved in.

Commercial organizations often want to increase the visibility of their web pages so that they can reach a wider audience in order to make their businesses successful. They are prepared to pay web search engines' owners to favor their web pages by displaying them at the top of search engine results. Organizations without the means to pay for such services are disadvantaged, therefore major web search engines, such as Google, provide alternative means of enhancing the visibility of web resources [15]. Search results obtained in such a manner are categorized as "organic".

SEO techniques are recommended for improved visibility of web resources in the organic search category. They are meant to enhance the indexing process of web pages by search engines [11–14]. Among others, if web page authors employ SEO techniques, this results in higher web page visibility in response to relevant user queries [16]. Elements that impact on the visibility of web pages are related to the metadata structure of the page, its content and the number of hyperlinks pointing to the web page in question [12–14,17,18]. SEO techniques "tune" the metadata structure and content of a web page to increase its visibility by web search engines. The number of hyperlinks and the way in which users refine their search query terms are beyond the web publisher's control [16]. However, the search behavior of users can be observed (or modeled) to understand how keywords (or key phrases) are employed when searching for information on a particular topic. This information can then be used to "tune" the metadata structure and content of a web page to further increase its visibility.

2.2. SEO Techniques in Academic Literature

There is currently little work in academic literature that discusses SEO techniques. The majority of information is available in non-academic platforms such as blogs, online discussion forums, websites and anecdotes [16]. The four main SEO techniques are keyword research, indexing, on-site optimization and off-site optimization [16]. On-site optimization consists of techniques applied to web pages in order to enhance their online visibility. For example, libraries employ on-site optimization techniques by duplicating online information about library resources as web pages that can be crawled and indexed to ultimately be searchable by web search engines; this resolves the invisibility of online library resources [13]. We follow a similar approach in this article by duplicating geoportal content (information about geospatial data) in web pages, which are discoverable by web search engines.

Two studies (part I and part II) [17,18] are among the few scientific publications about search engine optimization (SEO). They evaluated the impact of the content and metadata of a web page on the visibility by web search engines and found that not only the content of web pages, but also embedded metadata elements improve the visibility of web pages on the Web. After employing SEO techniques on web pages, [11] noticed an improvement in their visibility in search results with Google.

2.3. SEO Techniques with Schema.org and Dublin Core

The adoption of the Schema.org vocabulary had a major influence on SEO. Schema.org provides a generic set of vocabularies to annotate text content in HTML web pages, thereby infusing semantics into the web page, i.e., crawlers can "understand" a web page marked up with Schema.org. Major search engines, such as Google, Bing, Yahoo and Yandex, have endorsed Schema.org and support it [19]. Schema.org provides a number of themes (types) according to which the web page content can be categorized. Categories include products, people, organizations, places, events, cooking recipes, etc. The Schema.org vocabulary is embedded in the page content following certain encoding standards (mark up languages), such as microformat, RDFa or microdata. In this research, microdata is used.

Dublin Core provides a standard set of elements for describing digital resources on the Internet [20]. It has gained wide popularity due to its international and multidisciplinarity. Over the years, the Dublin Core Metadata Initiative (DCMI) has received criticism due to its inadequacy for describing complex resources and lack of detailed criteria in terms of how elements should be used in the context of local (custom) specific applications; geospatial applications are such an example. Much of the criticism hailed from the library community [21]. A criticism of concern to our study is that major web search engines do not consider Dublin Core metadata embedded in web pages [22]. Such an assertion has been neither explicitly confirmed nor denied by web search engine owners. For example, Google supports certain HTML meta tags elements such as title, description and charset [23]; these are also three Dublin Core elements.

Very few scientific studies have been conducted to test whether web search engines consider Dublin Core metadata elements. A study by [24] is an exception. They statistically compared the visibility of two sets of web pages to web search engines: a first set of pages without metadata and a second set with Dublin Core metadata elements. Results from seven web search engines, including Google and Yahoo, which is now powered by Bing, were evaluated. The statistical results showed that there was a difference in web search visibility between the two sets of pages: in five cases, pages with Dublin Core metadata outperformed those without metadata, for one search engine the difference was not significant, and for another the reverse was true: pages without metadata outperformed those with Dublin Core metadata. The study shows that indexing and retrieval algorithms of web search engines handle Dublin Core metadata differently. The Dublin Core vocabulary was included in this study to serve as baseline in the experiment for comparison with the Schema.org vocabulary, and also due to its popularity.

3. Method

3.1. Overview

In this section, we describe how the retrieval effectiveness of two metadata vocabularies (Dublin Core and Schema.org) with two web search engines (Google and Bing) was evaluated. Figure 1 provides an overview.

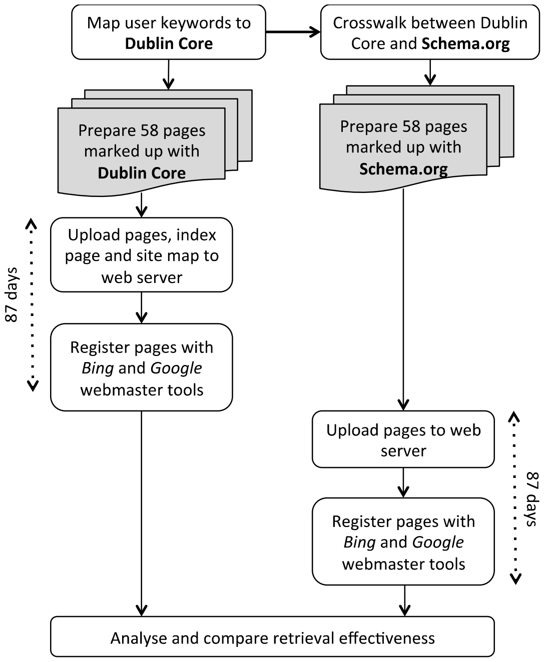
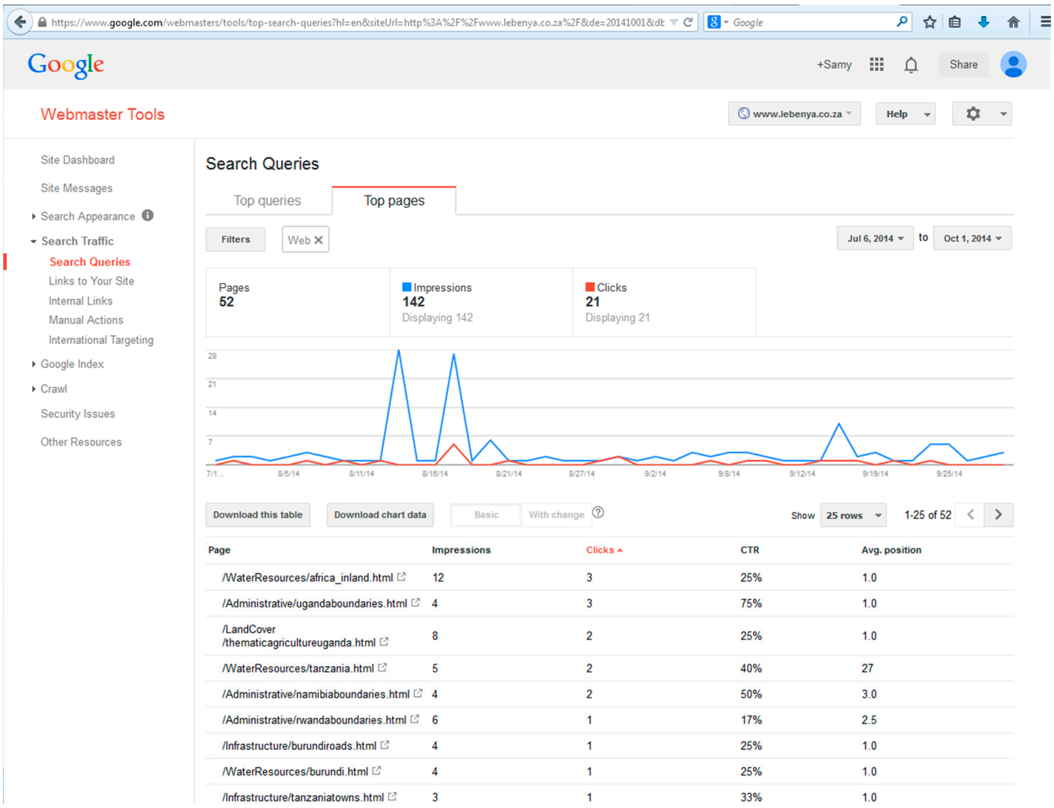

Map user keywords to **Dublin Core**
Crosswalk between Dublin Core and **Schema.org**
Prepare 58 pages marked up with **Dublin Core**
Prepare 58 pages marked up with **Schema.org**
Upload pages, index page and site map to web server
87 days
Register pages with *Bing* and *Google* webmaster tools
Upload pages to web server
Register pages with *Bing* and *Google* webmaster tools
87 days
Analyse and compare retrieval effectiveness

Figure 1. Overview of the study.

Two sets of 58 HTML web pages were prepared: one set marked up with elements of the Dublin Core metadata standard and the other with properties of types defined in the Schema.org vocabulary. The page content (geospatial metadata) in the two sets was identical. There was also an index page for each set, containing a direct link to each of the other 58 pages.

The first set of pages, marked up with Dublin Core, was uploaded to a web server exposed on a public domain (Internet) so that these pages could be crawled by search engine robots. The web server and the set of pages were registered with both Google and Bing via their respective webmaster tools. This ensures that search engines include the pages in their indexes, i.e., the pages are searchable. Furthermore, a sitemap (in XML format with a list of URL hyperlinks to the 58 pages in the set) was uploaded onto the web server and registered with each of the web search engines via their respective webmaster tools applications. The submission of sitemaps to web search engines is recommended as an SEO indexing technique because it enhances the crawling process (https://support.google.com/webmasters/answer/156184?hl=en), (http://www.bing.com/webmaster/help/how-to-submit-sitemaps-82a15bd4).

After a period of 87 days (6 July 2014 to 1 October 2014), the Dublin Core pages were replaced with the second set of pages, marked up with Schema.org, and these were then also registered with the Google and Bing webmaster tools. Figures 2 and 3 illustrate how the retrieval statistics for a page are displayed in the respective webmaster tools. For example, Figure 2 shows that africa_inland.html was included 12 times in Google search results with an average position of 1.0 (top of the list) in the list of results. In three cases the user clicked to see the page, i.e., a click through rate (CTR) of 25%. Figure 3 shows that mc_southsudan.html was included 12 times in Bing search results with an average position of 12.7 in the list of results. None of the users clicked to see the page, i.e., a CTR of 0%. To ensure that caching would not impact the results, each set of pages had unique names (e.g., Dublin Core file name: http://lebenya.co.za/Administrative/kenyaboundaries.html and corresponding Schema.org file name: http://lebenya.co.za/Administrative/mc_kenyaboundaries.html). Results for this second set of pages were also collected for a period of 87 days (23 November 2014 to 18 February 2015). According to literature about SEO techniques, a period of 87 days (i.e., 3 months) is adequate to allow the submitted pages to appear in the search results of web search engines (Google and Bing) [16].



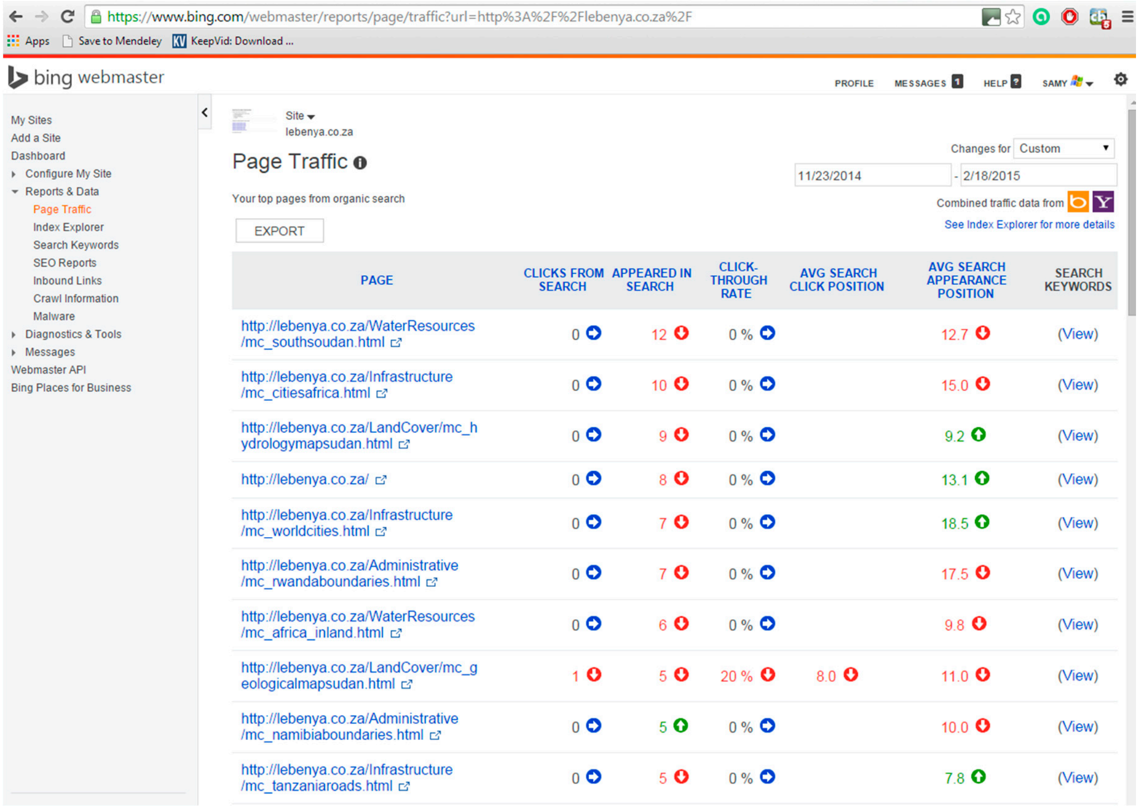
Search Queries

Jul 6, 2014 to Oct 1, 2014

Pages	🟦 Impressions	🟥 Clicks
52	142 Displaying 142	21 Displaying 21

Page	Impressions	Clicks ▲	CTR	Avg. position
/WaterResources/africa_inland.html	12	3	25%	1.0
/Administrative/ugandaboundaries.html	4	3	75%	1.0
/LandCover /thematicagricultureuganda.html	8	2	25%	1.0
/WaterResources/tanzania.html	5	2	40%	27
/Administrative/namibiaboundaries.html	4	2	50%	3.0
/Administrative/rwandaboundaries.html	6	1	17%	2.5
/Infrastructure/burundiroads.html	4	1	25%	1.0
/WaterResources/burundi.html	4	1	25%	1.0
/Infrastructure/tanzaniatowns.html	3	1	33%	1.0

Figure 2. Pages marked up with Dublin Core listed on Google Webmaster Tools.



https://www.bing.com/webmaster/reports/page/traffic?url=http%3A%2F%2Flebenya.co.za%2F

lebenya.co.za

Page Traffic

Your top pages from organic search

Changes for Custom

11/23/2014 - 2/18/2015

Combined traffic data from

See Index Explorer for more details

PAGE	CLICKS FROM SEARCH	APPEARED IN SEARCH	CLICK-THROUGH RATE	AVG SEARCH CLICK POSITION	AVG SEARCH APPEARANCE POSITION	SEARCH KEYWORDS
http://lebenya.co.za/WaterResources/mc_southsoudan.html	0 →	12 ↓	0 % →		12.7 ↓	(View)
http://lebenya.co.za/Infrastructure/mc_citiesafrica.html	0 →	10 ↓	0 % →		15.0 ↓	(View)
http://lebenya.co.za/LandCover/mc_hydrologymapsudan.html	0 →	9 ↓	0 % →		9.2 ↑	(View)
http://lebenya.co.za/	0 →	8 ↓	0 % →		13.1 ↑	(View)
http://lebenya.co.za/Infrastructure/mc_worldcities.html	0 →	7 ↓	0 % →		18.5 ↑	(View)
http://lebenya.co.za/Administrative/mc_rwandaboundaries.html	0 →	7 ↓	0 % →		17.5 ↓	(View)
http://lebenya.co.za/WaterResources/mc_africa_inland.html	0 →	6 ↓	0 % →		9.8 ↓	(View)
http://lebenya.co.za/LandCover/mc_geologicalmapsudan.html	1 ↓	5 ↓	20 % ↓	8.0 ↓	11.0 ↓	(View)
http://lebenya.co.za/Administrative/mc_namibiaboundaries.html	0 →	5 ↑	0 % →		10.0 ↓	(View)
http://lebenya.co.za/Infrastructure/mc_tanzaniaroads.html	0 →	5 ↓	0 % →		7.8 ↑	(View)

Figure 3. Pages marked up with Schema.org listed on Bing Webmaster.

The webmaster tools provided reports with information about page traffic, crawl information and search keywords (queries) that triggered appearances of the pages in the search results of respective search engines, the number of pages that were retrieved, the number of times a page appeared in a search engine result (impressions) and the average position (rank) of a page in the list of search engine results. The number of pages that were retrieved and rank statistics were analyzed to evaluate retrieval effectiveness of the web search engines. The retrieval effectiveness of four combinations of vocabularies and search engine results were analyzed and compared:

- Dublin Core: Bing vs. Google
- Schema.org: Bing vs. Google
- Bing: Dublin Core vs. Schema.org
- Google: Dublin Core vs. Schema.org

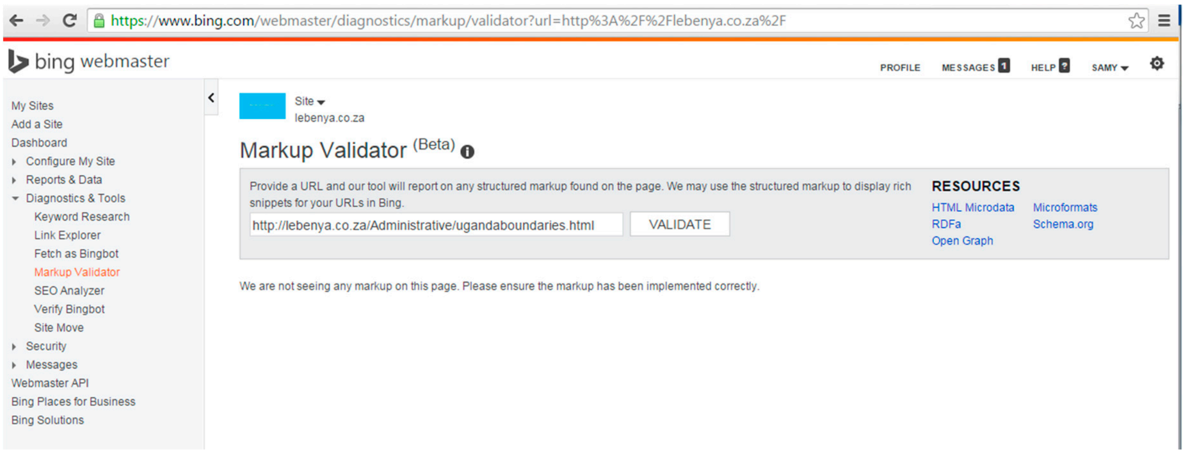
The "tuning" of the content of the page and registration of these pages with the search engines represents the three SEO techniques that were used: keyword research informed by the taxonomy of observed user search terms; on-site optimization was done by infusing metadata into the pages based on the taxonomy; and indexing was initiated by registering the pages with the search engines and submitting site maps to the search engines.

3.2. Justification for Using Bing, Google, Dublin Core and Schema.org

Dublin Core is intended to facilitate "cross-discipline discovery" and assuring interoperability across various disciplines of interest [25]. Its elements can be encoded using RDF/XML, XML and HTML "meta" tag. The simplicity and ease of use, combined with the generic nature of terms (elements) employed in the Dublin Core standard, make it appropriate for this study. Schema.org which is an initiative endorsed by Google, Bing, Yahoo and Yandex search engines [19], provides a shared vocabulary that is generic enough to accommodate popular concepts. It aims to enable machines (web search engines) to understand the contents of web resources (web pages) thereby enhancing their online discoverability. Google and Bing web search engines were considered in this study based on their popularity, which is well documented in the literature and they have both endorsed Schema.org [19].

Because the retrieval effectiveness of the marked up pages could be influenced by how the web search engines (Bing and Google) recognize and interpret the vocabularies (Dublin Core and Schema.org), we verified if and how each search engine recognized the two kinds of vocabularies. The "Markup Validator" in Bing webmaster tools and the "Structured Data Testing Tool" of Google webmaster tools were employed.

At the time of testing, Bing did not identify any tags in the pages marked up with Dublin Core vocabulary using the "meta" tag. The results of the "Markup Validator" in Bing in Figure 4 show that Dublin Core tags were not recognized for ugandaboundaries.html ("We are not seeing any markup on this page"). However, the selected types of Schema.org vocabulary marked up with microdata tags were recognized by the "Markup Validator" in Bing. Figure 5 shows the tags that were recognized, namely, ItemPage, Person, Place, GeoCoordinates, etc. In contrast, the Google "Structured Data Testing Tool" identified tags in the set of Dublin Core pages. Figure 6 shows the results of the Structured Data Testing Tool" of Google for ugandaboundaries.html. Tags that were recognized are listed under "Custom Search Result Filters". The selected types of Schema.org vocabulary using microdata were also recognized. Figure 7 illustrates the types that were recognized in ugandaboundaries.html, namely, author, creator, content, provider, etc. The capabilities of the search engines to recognize and interpret the vocabularies were considered when the results were analyzed.

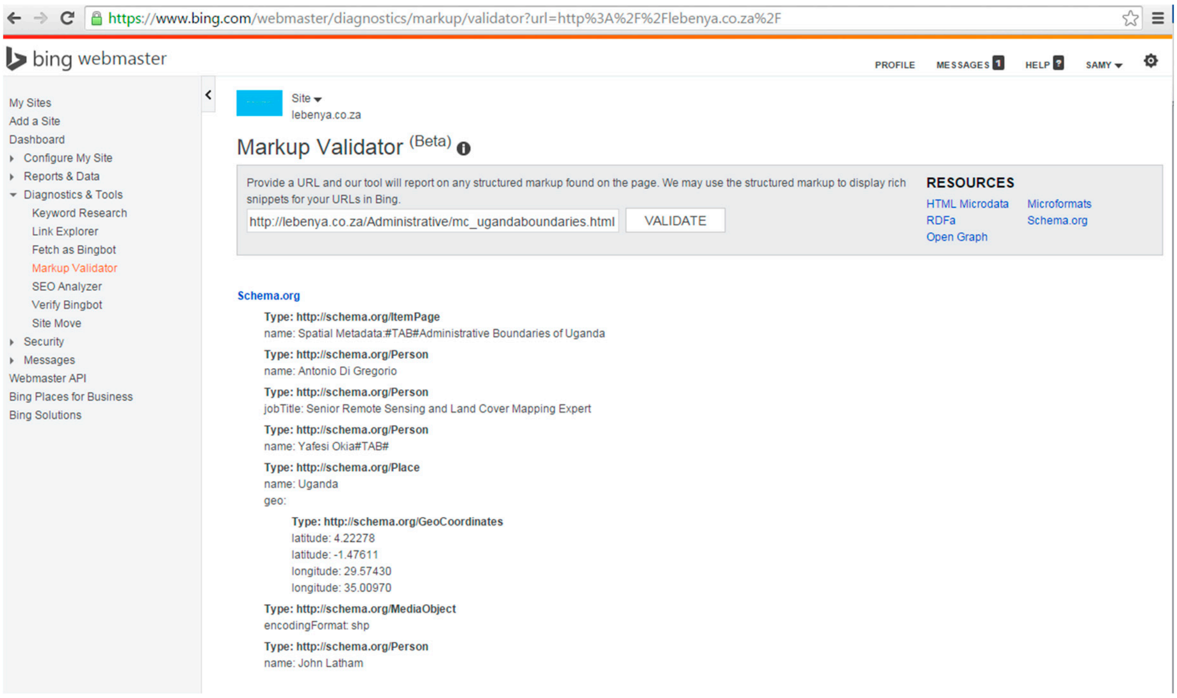


Provide a URL and our tool will report on any structured markup found on the page. We may use the structured markup to display rich snippets for your URLs in Bing.

http://lebenya.co.za/Administrative/ugandaboundaries.html

We are not seeing any markup on this page. Please ensure the markup has been implemented correctly.

Figure 4. Bing "Markup Validator": Result for ugandaboundaries.html, marked up with Dublin.



Markup Validator (Beta)

Provide a URL and our tool will report on any structured markup found on the page. We may use the structured markup to display rich snippets for your URLs in Bing.

http://lebenya.co.za/Administrative/mc_ugandaboundaries.html

Schema.org

Type: http://schema.org/ItemPage
name: Spatial Metadata:#TAB#Administrative Boundaries of Uganda

Type: http://schema.org/Person
name: Antonio Di Gregorio

Type: http://schema.org/Person
jobTitle: Senior Remote Sensing and Land Cover Mapping Expert

Type: http://schema.org/Person
name: Yafesi Okia#TAB#

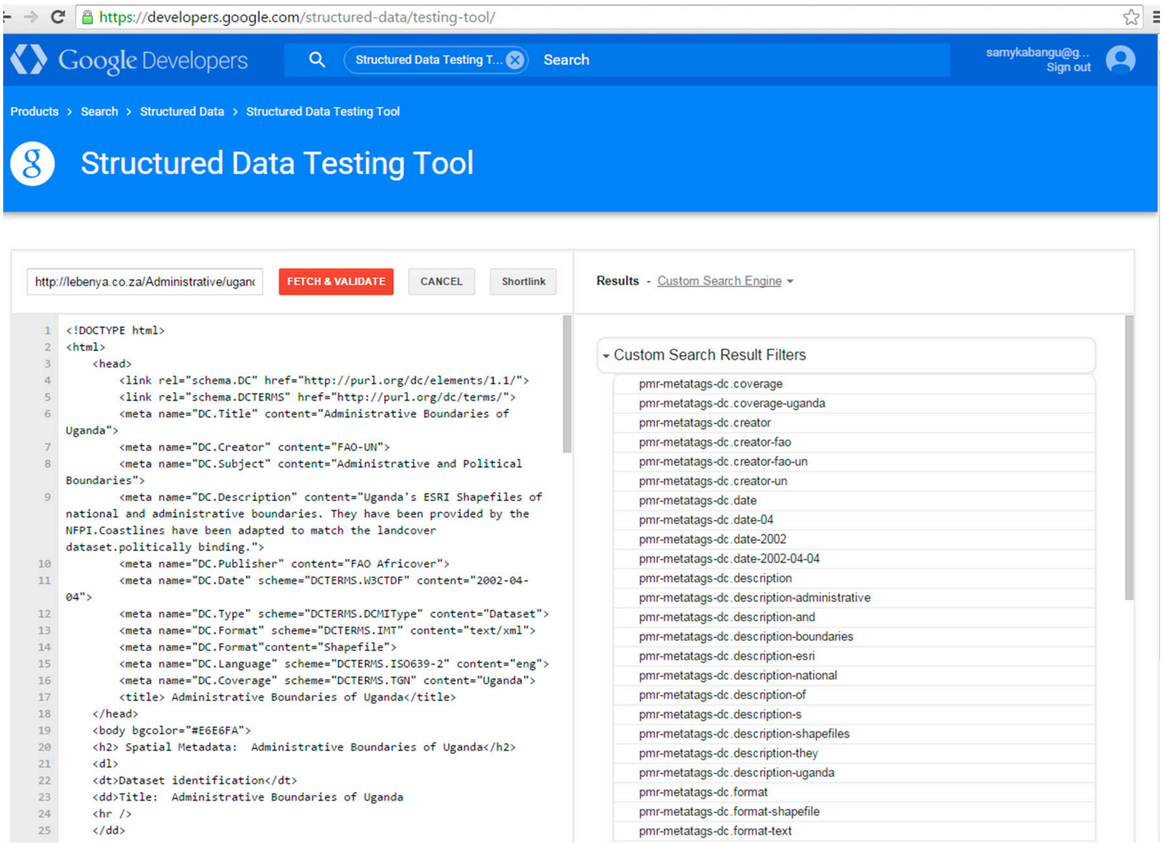
Type: http://schema.org/Place
name: Uganda
geo:

Type: http://schema.org/GeoCoordinates
latitude: 4.22278
latitude: -1.47611
longitude: 29.57430
longitude: 35.00970

Type: http://schema.org/MediaObject
encodingFormat: shp

Type: http://schema.org/Person
name: John Latham

Figure 5. Bing Markup Validator: Result for ugandaboundaries.html marked up with Schema.org using microdata.


https://developers.google.com/structured-data/testing-tool/
Google Developers
Structured Data Testing T...
Search
samykabangu@g...
Sign out
Products
Search
Structured Data
Structured Data Testing Tool
FETCH & VALIDATE
CANCEL
Shortlink
Results - Custom Search Engine

Structured Data Testing Tool

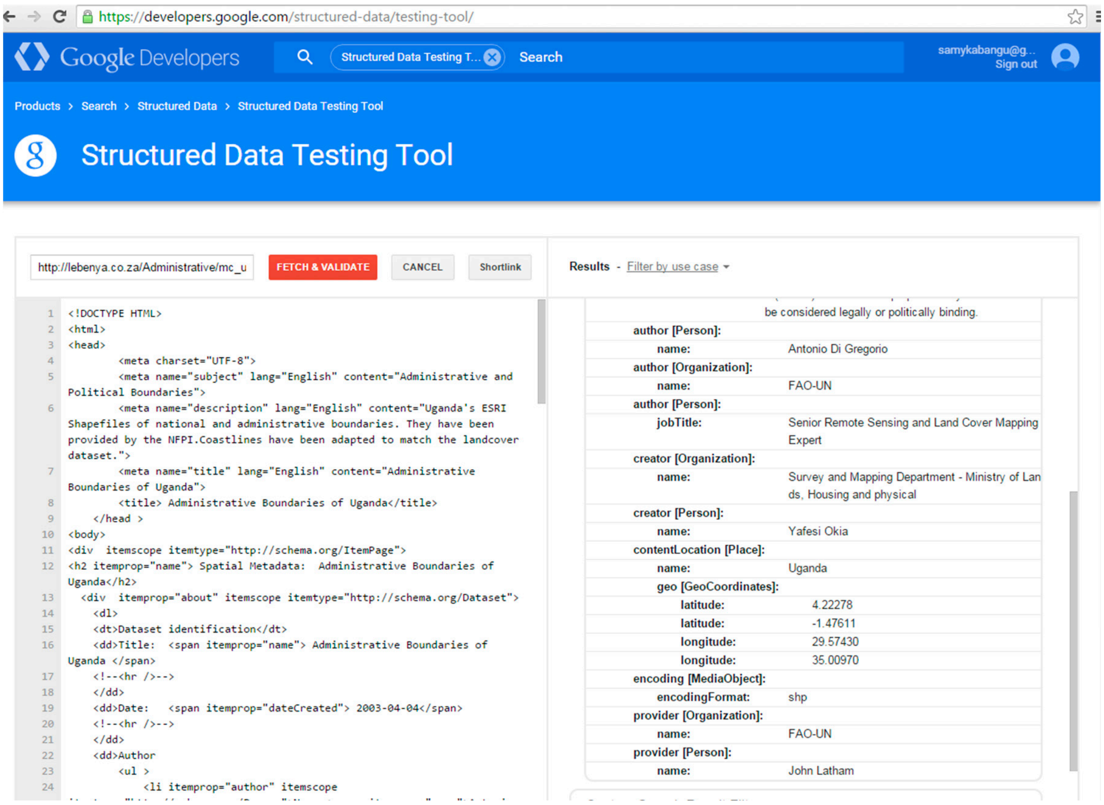
http://lebenya.co.za/Administrative/uganc

```
<!DOCTYPE html>
<html>
    <head>
        <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
        <link rel="schema.DCTERMS" href="http://purl.org/dc/terms/">
        <meta name="DC.Title" content="Administrative Boundaries of
Uganda">
        <meta name="DC.Creator" content="FAO-UN">
        <meta name="DC.Subject" content="Administrative and Political
Boundaries">
        <meta name="DC.Description" content="Uganda's ESRI Shapefiles of
national and administrative boundaries. They have been provided by the
NFPI.Coastlines have been adapted to match the landcover
dataset.politically binding.">
        <meta name="DC.Publisher" content="FAO Africover">
        <meta name="DC.Date" scheme="DCTERMS.W3CTDF" content="2002-04-
04">
        <meta name="DC.Type" scheme="DCTERMS.DCMIType" content="Dataset">
        <meta name="DC.Format" scheme="DCTERMS.IMT" content="text/xml">
        <meta name="DC.Format"content="Shapefile">
        <meta name="DC.Language" scheme="DCTERMS.ISO639-2" content="eng">
        <meta name="DC.Coverage" scheme="DCTERMS.TGN" content="Uganda">
        <title> Administrative Boundaries of Uganda</title>
    </head>
    <body bgcolor="#E6E6FA">
    <h2> Spatial Metadata:  Administrative Boundaries of Uganda</h2>
    <dl>
    <dt>Dataset identification</dt>
    <dd>Title:  Administrative Boundaries of Uganda
    <hr />
    </dd>
```

Custom Search Result Filters

- pmr-metatags-dc.coverage
- pmr-metatags-dc.coverage-uganda
- pmr-metatags-dc.creator
- pmr-metatags-dc.creator-fao
- pmr-metatags-dc.creator-fao-un
- pmr-metatags-dc.creator-un
- pmr-metatags-dc.date
- pmr-metatags-dc.date-04
- pmr-metatags-dc.date-2002
- pmr-metatags-dc.date-2002-04-04
- pmr-metatags-dc.description
- pmr-metatags-dc.description-administrative
- pmr-metatags-dc.description-and
- pmr-metatags-dc.description-boundaries
- pmr-metatags-dc.description-esri
- pmr-metatags-dc.description-national
- pmr-metatags-dc.description-of
- pmr-metatags-dc.description-s
- pmr-metatags-dc.description-shapefiles
- pmr-metatags-dc.description-they
- pmr-metatags-dc.description-uganda
- pmr-metatags-dc.format
- pmr-metatags-dc.format-shapefile
- pmr-metatags-dc.format-text

Figure 6. Google "Structured data testing tool": Result for ugandaboundaries.html, marked up with Dublin Core.


https://developers.google.com/structured-data/testing-tool/
Google Developers
Structured Data Testing T...
Search
samykabangu@g...
Sign out
Products > Search > Structured Data > Structured Data Testing Tool
Structured Data Testing Tool
http://lebenya.co.za/Administrative/mc_u
FETCH & VALIDATE
CANCEL
Shortlink
Results - Filter by use case

```
<!DOCTYPE HTML>
<html>
<head>
        <meta charset="UTF-8">
        <meta name="subject" lang="English" content="Administrative and
Political Boundaries">
        <meta name="description" lang="English" content="Uganda's ESRI
Shapefiles of national and administrative boundaries. They have been
provided by the NFPI.Coastlines have been adapted to match the landcover
dataset.">
        <meta name="title" lang="English" content="Administrative
Boundaries of Uganda">
        <title> Administrative Boundaries of Uganda</title>
    </head >
<body>
<div  itemscope itemtype="http://schema.org/ItemPage">
<h2 itemprop="name"> Spatial Metadata:  Administrative Boundaries of
Uganda</h2>
  <div  itemprop="about" itemscope itemtype="http://schema.org/Dataset">
    <dl>
    <dt>Dataset identification</dt>
    <dd>Title:  <span itemprop="name"> Administrative Boundaries of
Uganda </span>
    <!--<hr />-->
    </dd>
    <dd>Date:   <span itemprop="dateCreated"> 2003-04-04</span>
    <!--<hr />-->
    </dd>
    <dd>Author
        <ul >
            <li itemprop="author" itemscope
```

```
                              be considered legally or politically binding.
    author [Person]:
        name:                 Antonio Di Gregorio
    author [Organization]:
        name:                 FAO-UN
    author [Person]:
        jobTitle:             Senior Remote Sensing and Land Cover Mapping
                              Expert
    creator [Organization]:
        name:                 Survey and Mapping Department - Ministry of Lan
                              ds, Housing and physical
    creator [Person]:
        name:                 Yafesi Okia
    contentLocation [Place]:
        name:                 Uganda
        geo [GeoCoordinates]:
            latitude:             4.22278
            latitude:             -1.47611
            longitude:            29.57430
            longitude:            35.00970
    encoding [MediaObject]:
        encodingFormat:       shp
    provider [Organization]:
        name:                 FAO-UN
    provider [Person]:
        name:                 John Latham
```

Figure 7. Google "Structured data testing tool": Result for ugandaboundaries.html, marked up with Schema.org using microdata.

3.3. *Method for Marking up Geospatial Metadata*

Metadata for selected datasets available on the "FAO Africover" geoportal website (http://www.fao.org/geonetwork/srv/en/main.home) was adapted and infused into the web pages according to a taxonomy of observed user search terms. The taxonomy was constructed by analysing and assigning search terms (i.e., a string of alphanumeric characters entered by a participant or web searcher, e.g., "namibia" or "shapefile") obtained from the Bing Webmaster tools log and from a user experiment into four categories (Application Domain, Location, Feature Type and Data Model) [26]. The taxonomy represents the terms employed by users when searching for geospatial data on the web using general-purpose web search engines.

Categories in the taxonomy were mapped to equivalent elements in the ISO metadata standard for geographic information (ISO 19115:2003). Next, ISO 19115:2003 elements were mapped to the Dublin Core vocabulary according to the withdrawn CWA 14857:2003 (E) standard. The geospatial metadata in the HTML web pages was marked up with the Dublin Core metadata vocabulary. A mapping (crosswalk) between the Dublin Core vocabulary and Schema.org was created, following the methodology described in [27]. The crosswalk informed the construction of pages in which geospatial metadata was marked up with Schema.org vocabulary using microdata. Figure 8 illustrates the chain of mapping from the taxonomy of search terms through to Schema.org.

Table 1 shows an example of such a mapping. For example, "Subject (Theme)" represents a category in the taxonomy and "Road Network" is an example of a value in that category. The corresponding value for the "TITLE" element (or Term) of the Dublin Core vocabulary is "Roads of Egypt", while the corresponding "NAME" element of Schema.org also contains the value "Road Egypt". Table 1 also illustrates the marking up in the two vocabularies. Refer to Appendix A for a high level description of the mapping (crosswalk) from taxonomy via ISO 19115:2003 spatial metadata standard to Dublin Core and through to Schema.org.

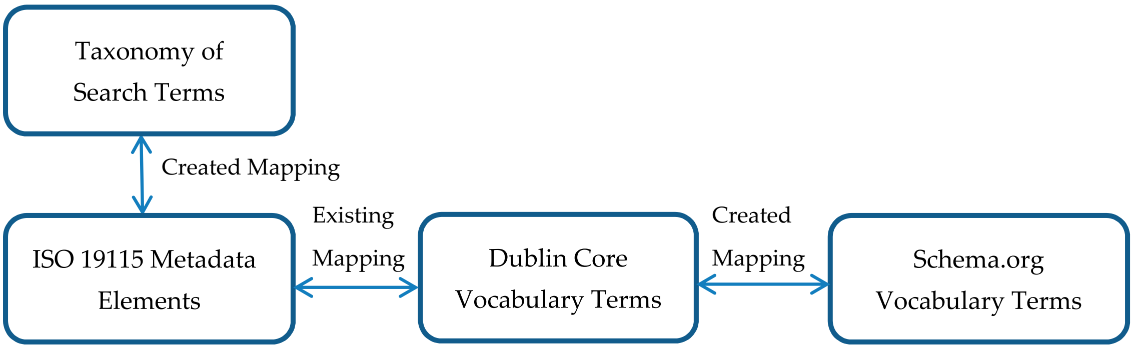

Taxonomy of Search Terms
Created Mapping
ISO 19115 Metadata Elements
Existing Mapping
Dublin Core Vocabulary Terms
Created Mapping
Schema.org Vocabulary Terms

Figure 8. Mapping from taxonomy to vocabularies via standards.

Table 1. Example mapping from taxonomy to Schema.org.

Taxonomy	**Subject (Theme):** ***Road Network***
ISO 19115	**Dataset title**: *Roads of Egypt*
Dublin Core Vocabulary	**TITLE**: *Roads of Egypt* <head> <link rel = "schema.DC" href = http://purl.org/dc/elements/1.1> <link rel = "schema.DCTERMS" href =http://purl.org/dc/terms"/> <meta name = "DC.Title" content = *"Roads of Egypt" lang = "en"*> . . . </head>
Schema.org Vocabulary	**NAME**: *Roads of Egypt* <body> <div itemscope itemtype="http://schema.org/ItemPage"> <h2 itemprop="name"> Roads of Egypt</h2> . . . </div> </body>

3.4. Method for the Evaluation of Retrieval Effectiveness

Ali, R., et al. [28] reviewed different methods for evaluating the retrieval effectiveness of web search engines and grouped them into the following eight categories: relevance based evaluation, ranking based evaluation, user satisfaction based evaluation, size/coverage of web-based evaluation, dynamics of search results based evaluation, few relevant/known item based evaluation, specific topic/domain based evaluation and automatic evaluation. In this research, we apply "relevance based evaluation" and "ranking based evaluation". We evaluated whether relevant pages are included in search results, and if so, we analyzed the number of relevant pages that were retrieved, and the average ranking positions of these retrieved pages in the search results. Although not included in the analysis, we also recorded the frequency of inclusion of the retrieved pages in search results (also known as impressions).

Interpreting the definitions by [29], in this study recall refers to the ratio of relevant retrieved pages to all relevant pages. In most experiment settings reported in the literature [28–30], it is difficult to determine the total number of relevant pages to be retrieved by web search engines. However, in our case, a set of 58 pages was deemed relevant and was used to calculate "recall". Precision, the ratio of relevant retrieved pages to all retrieved pages (number of pages in the search result) is an alternative measure of relevance. Given the size of the web, a significantly high number of pages (in the order of millions) are likely to be retrieved in response to a given query, resulting in a precision value that is a very small fraction, which does not add much value to the analysis. For this reason, the precision measure is not included in the results.

In controlled lab experiments, a defined set of queries is used to measure recall and precision. However, this experiment was intentionally user-driven, i.e., the queries were not pre-defined by the researchers but rather formulated by "real" users looking for geospatial data on the Web. This approach was followed because it is already known that using a defined set of queries will result in the relevant pages being included in the search results (see e.g., [17,18]). For this research, we wanted to test whether the taxonomy of user search terms (which reflects our understanding of how users are likely to formulate their queries, see description in Section 3.3) would render the pages "visible" to "real" user queries. This experiment design reflects the real world where one does not know who the users are and how they formulate their queries. User queries that triggered the appearance of the submitted pages in search results of one of the search engines are listed in Appendix B.

"Recall", was used as a measure of relevance in the Bing and Google search results. Ranking was quantified in terms of the average position of a page in the search results; the lower the average position, the higher the ranking. Ranking was employed in the "ranking based evaluation"; it is also a measure of page visibility and by implication geospatial resource discoverability. Descriptions of the ranking-based evaluation measure ("Rank") and relevance-based evaluation measure ("Recall") are provided in Table 2.

Table 2. Measures used to evaluate retrieval effectiveness.

Measure	Description
Rank	The average position of a page (URL) in the search results. The lower the average position, the higher its rank and visibility.
Recall	The ratio of the number of relevant pages retrieved (included in the search results) to the total number of relevant pages in the web server (58 for this research).

In order to determine whether there was a statistical difference in the ranking of the two sets of results (that is, pages marked up with Dublin Core and those marked up with Schema.org), the Wilcoxon Signed Rank test was employed. Statistical significance determines whether the difference in the ranking positions of the sets of results is not a random occurrence. Various studies have used statistical significance to compare retrieval effectiveness of web search engines [28]. In a specific example, for their comparison of automatic and manual (by humans) methods for determining the relevance of web search results, Can, F., et al. [30] performed a number of statistical significance tests based on precision and recall measures obtained from eight different web search engines. Zhang, J., et al. [18] employed three statistical techniques (one-way ANOVA, two-way ANOVA, and independent-sample *t*-test) to test whether there was a statistically significant difference in the ranking positions of 46 web pages among 19 web search engines.

In this study, given the fact that the results (average position) from the two samples are dependent (related) and paired, the Wilcoxon Signed Rank test was chosen to test whether there is a statistically significant difference in the average positions of web pages in the search results of the two web search engines. Because some pages never appeared in any search results, they were assigned an average position of 1000 so that one-to-one pairing between pages could be done. Hence, there was no need to test for normality.

4. Results: Retrieval Effectiveness Evaluation

4.1. Relevance-Based Evaluation

As illustrated in Table 3, 49 of the 58 pages marked up with Dublin Core appeared in the search results of Google and 15 pages appeared in the search results of Bing. This amounts to a recall of 0.84 for Google, and 0.26 for Bing. Hence, it can be said that relevance-based retrieval effectiveness of pages marked up with Dublin Core is higher with Google than with Bing. This result can be explained by the fact that Bing did not recognize the Dublin Core mark up. Another explanation could be that Google achieved a higher recall because it is used by more people (if more searches are done, the probability of more impressions is higher).

Again, as illustrated in Table 3, 45 of the 58 pages marked up with Schema.org appeared in the search results of Google and only 39 in the search results of Bing. This amounts to a recall of 0.78 for Google and 0.67 for Bing. Similar to the observation above, relevance-based retrieval effectiveness of pages marked up with Schema.org is higher with Google than with Bing, despite both search engines detecting the Schema.org vocabulary.

Table 3. Results: Relevance based retrieval effectiveness.

		Bing	**Google**
Dublin Core	Number of retrieved relevant pages	15	49
	Number of relevant pages	58	58
	Recall	0.26	0.84
Schema.org	Number of retrieved relevant pages	39	45
	Number of relevant pages	58	58
	Recall	0.67	0.78

When comparing the relevance-based retrieval effectiveness of the two vocabularies, it can be seen that in Bing, pages marked up with Dublin Core had lower recall values than those marked up with Schema.org (0.26 vs. 0.67). In contrast, in Google, pages marked up with Dublin Core achieved a higher recall than those marked-up with Schema.org (0.84 vs. 0.78). Once again, this result can be explained by the fact that Bing did not recognize the Dublin Core mark up.

For both search engines and both vocabularies, some pages never appeared in any of the search results.

4.2. Ranking-Based Evaluation

The descriptive statistics in Table 4 show that for both vocabularies the ranking is generally higher (i.e., lower average position) in Google than in Bing. The difference is more significant for Dublin Core, e.g., a mean average position of 163.1 (Google) vs. 744.5 (Bing) and a mean average position of 227.1 (Google) vs. 338.2 (Bing) with Schema.org. In Google, the ranking is generally higher (i.e., lower average position) for pages marked up with Dublin Core (163.1 vs. 227.1); the opposite was observed for Bing: the ranking is generally higher for pages marked up with Schema.org (744.5 vs. 338.2). The latter can again be explained by the fact that Bing did not recognize the Dublin Core mark up.

Table 4. Results: Ranking based retrieval effectiveness.

	Rank	**Bing**	**Google**
Dublin Core	Mean average position	744.5	163.1
	Standard deviation of the average position	436.4	363.5
	Minimum position	1	1
	Maximum position *	1000	1000
Schema.org	Mean average position	338.2	227.1
	Standard deviation of the average position	466.4	419.1
	Minimum position	2	1
	Maximum position *	1000	1000

* A max value of 1000 assigned to pages that did not appear in the search results.

Table 5 shows the results of the pairwise comparison between the ranks of pages retrieved by the two search engines. For 47 pages marked up with Dublin Core, the rankings in Google search results were higher (i.e., lower average position) than in Bing search results. For 4 pages the reverse was true. For 7 pages, the rankings in Google and Bing search results were equal. For 42 pages marked up with

Schema.org, the rankings in Google search results were higher (i.e., lower average position) than in Bing search results. For 11 pages the reverse was true. For 5 pages, the rankings in Google and Bing search results were equal.

Table 5. Results: Pairwise comparison of rankings.

	Bing vs. Google	n	Mean Rank	Sum of Ranks
Dublin Core	Bing rank < Google rank	47	27.06	1272
	Bing rank > Google rank	4	13.50	54
	Bing rank = Google rank	7		
		58		
Schema.org	Bing rank < Google rank	42	25.98	1091
	Bing rank > Google rank	11	30.91	340
	Bing rank = Google rank	5		
		58		

For the Wilcoxon Signed Rank test, the null hypothesis was set to be that the average positions between the two samples (Bing and Google results) are equal. For both Dublin Core and Schema.org, the results (with p-value = 0.001) of the Wilcoxon Signed Rank tests performed in the SPSS 24 statistical software package (https://www-01.ibm.com/software/) suggest that at 5% level of significance the null hypothesis should be rejected. This was confirmed in both cases by a p-value (0.001) smaller than the significance level (0.05). It can therefore be said that pages marked up with either Dublin Core or Schema.org were ranked higher on Google than on Bing, i.e., ranking-based retrieval effectiveness is higher on Google than Bing for both vocabularies.

Table 6 shows the results of the page-by-page ranking comparison. For Bing, 35 pages marked up with Schema.org were ranked higher in the results (i.e., lower average position) than the same pages marked up with Dublin Core. The reverse was true for 8 pages. 15 pages were ranked equally in either vocabulary. For Google, 12 pages marked up with Schema.org were ranked higher (i.e., lower average position) than Bing. The reverse was true for 15 cases. 31 pages were ranked equally in either vocabulary.

Table 6. Results: Page-by-page comparison of ranking.

	Dublin Core vs. Scheman.org	n	Mean Rank	Sum of Ranks
Bing	Dublin Core rank < Schema.org rank	35	21.59	755.50
	Dublin Core rank > Schema.org rank	8	23.81	190.50
	Dublin Core rank = Schema.org rank	15		
		58		
Google	Dublin Core rank < Schema.org rank	12	13.25	159
	Dublin Core rank > Schema.org rank	15	14.6	219
	Dublin Core rank = Schema.org rank	31		
		58		

Again, the null hypothesis was set to be that the average positions between the two samples (Dublin Core and Schema.org pages) are equal. The results from the Wilcoxon Signed Rank tests performed in the SPSS 24 statistical software package on the Bing rankings suggest that at a 5% level of significance the null hypothesis should be rejected. This is confirmed by the p-value (0.001) being smaller than the significance level (0.05). Hence, it can be concluded that pages marked up with Schema.org were ranked higher (i.e., had higher visibility) in the Bing search results than pages marked up with Dublin Core.

The results from the Wilcoxon Signed Rank tests for the Google rankings suggest that at a 5% level of significance, the null hypothesis be accepted. This is confirmed by the *p*-value (0.458) being larger than the significance level (0.05). Hence, it can be concluded that pages marked up with Schema.org vocabulary achieved the same level of visibility (ranking or average position) in the Google search results as those marked up with Dublin Core.

4.3. Summary of Results

An overview of the retrieval effectiveness evaluation is presented in Table 7. These results are discussed in the subsequent section.

Table 7. Summary of results: Relevance-based and ranking-based retrieval effectiveness.

Relevance-based evaluation of retrieval effectiveness	Vocabulary	Dublin Core	Google recall > Bing recall
		Schema.org	Google recall > Bing recall
	Search engine	Bing	Schema.org recall > Dublin Core recall
		Google	Dublin Core recall > Schema.org
Ranking-based evaluation of retrieval effectiveness	Vocabulary	Dublin Core	Google visibility > Bing visibility
		Schema.org	Google visibility > Bing visibility
	Search engine	Bing	Schema.org visibility > Dublin Core visibility
		Google	Schema.org visibility = Dublin Core visibility

5. Discussion and Conclusions

The research presented in this paper demonstrates, through an empirical study, how SEO techniques can be leveraged to improve the visibility and discoverability of geospatial resources on the Web. The research emanates from the need to find solutions for current impediments to the discovery of geospatial resources in geoportals.

Search terms typically employed by users looking for geospatial resources on the Web were categorized and classified into a taxonomy. A chain of mapping from the taxonomy to ISO 19115:2003, *Geographic information—Metadata*, to the Dublin Core vocabulary and finally to the Schema.org vocabulary, guided the insertion of geospatial metadata into the pages. By mapping search terms employed by general users (in the taxonomy) to terminology defined by experts (in vocabularies and standards), the discovery of geospatial resources is enhanced for mainstream information retrieval by general-purpose web search engines.

The retrieval effectiveness of web pages marked up with Dublin Core was evaluated and compared to that of pages marked up with Schema.org. Overall, the results indicate that Google was more effective in retrieving the pages than Bing; and pages marked up with Schema.org were more effectively retrieved than those marked up with Dublin Core. The latter difference was found to be statistically insignificant for Google, suggesting that there is no difference between Dublin Core and Schema.org with the Google search engine.

In related work, CSW-type geospatial metadata hidden behind geoportals was exposed in linked data format for crawling and indexing by search engines [5]. The research reported in this paper takes this work a step further by performing empirical tests to ascertain that this approach improves visibility of the geospatial resources, and by demonstrating how user search terms can assist in improving the visibility. SEO techniques were employed for this. Lopez-Pellicer, J.F., et al. [5] used RDF to establish links between metadata concepts in one ontology to related concepts in other ontologies. This facilitates machine understanding of such concepts. We decided to use Schema.org for assigning meaning to the contents of the HTML page, and the markup was done in Microdata, the mark up language of choice for Schema.org.

The research demonstrated how search terms can be identified and used to infuse metadata according to the Dublin Core and Schema.org vocabularies into web pages; and how, along with the registration of the pages for indexing, the discovery of geospatial resources by general purpose web search engines can be enhanced. In contrast, others have proposed and developed special-purpose web crawlers that discover and understand geospatial resources in a geoportal [6,7], but this approach does not expose the resources for mainstream information retrieval. The research also contributes to online information retrieval generally by demonstrating the application of SEO techniques in the geospatial domain. The retrieval effectiveness evaluation presented in this paper helps to understand if and how search engines detect and index geospatial metadata embedded in web pages.

The approach presented in this paper comes not as a way of replacing geoportals, but rather as a means of complementing geoportals to achieve their intended goal. Geoportals of today are mainly built on OGC's CSW [3], which was not designed to be crawled by web search engines [5], i.e., any CSW-accessible metadata is not indexed by search engines. Publishing geospatial metadata in HTML pages leads to increased web traffic to geoportals where the relevant geospatial data can be further explored, accessed or downloaded. This can be described with an analogy of an iceberg: geoportals of today are massive icebergs concealed below the surface of the sea of web resources. The approach described in this paper reveals the tip of the iceberg for further exploration underneath the surface.

During the construction of the taxonomy for this research, search terms were manually allocated to categories based on subjective judgment. Future work could explore automatic categorization, assisted by human supervision. Furthermore, given that only descriptive statistics and inferential statistics were performed on small samples of web pages and search queries, advanced inferential statistical analyses could be performed on bigger sets of pages. Possibilities for broadening the taxonomy to have more levels, so as to include more search terms, should also be explored.

In future work, ontology-based semantic interoperability could be considered for the mapping. That is, besides Schema.org and Dublin Core, additional ontologies could be considered and evaluated for including geospatial metadata in the web pages. Also, search engine results should be collected for extended periods to determine whether this has any effect on the retrieval effectiveness. Also, this research considered the overall results, rather than the detailed dynamics of the results for a single query or a set of queries. The detailed results could be investigated in future work, e.g., by measuring Recall@N and Precision@N for a specific query or a set of queries, by varying N (N is a cut off for the highest ranked pages in the results). This research considered relevance based evaluation and ranking based evaluation. Experiments with other retrieval effectiveness evaluation methods, such as user satisfaction based evaluation, size/coverage of web-based evaluation, dynamics of search results based evaluation, few relevant/known item based evaluation, specific topic/domain based evaluation and automatic evaluation, could also be conducted.

Acknowledgments: The authors would like to thank Professor Lopez-Pellicer (University of Zaragoza, Spain) and Professor Francois Steffens (University of Pretoria, South Africa) for the suggestions and guidance they provided with regards to the choice of appropriate statistical analysis tests to perform in this research.

Author Contributions: The research was conducted by S.K. for a Master of Science (MSc) degree in Geoinformatics under the supervision of S.C.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. High level mapping (crosswalk) description.

Taxonomy of Search Terms Categories	Corresponding ISO 19115 Elements (Adapted from the Long Name)	Corresponding Dublin Core Elements	Corresponding Schema.org Properties (Elements)
Subject (Theme)	Title (Name by which the cited resource is known)	**TITLE** (A name given to the resource. Typically, a Title will be a name by which the resource is formally known)	**NAME** (The name of the item)
	Topic Category (Main theme(s) of the dataset	**SUBJECT** (The topic of the content of the resource. Typically, a Subject will be expressed as keywords, key phrases, or classification codes that describe a topic of the resource)	**ABOUT** (The subject matter of the content)
	Abstract (Brief narrative summary of the content of the resource(s)	**DESCRIPTION** (An account of the resource. Description may include but is not limited to; an abstract, table of contents, reference to graphical representation of content or a free-text account of the content)	**DESCRIPTION** (A description of the item)
	Lineage (Detailed description of the level of the source data)	**SOURCE** (A reference to a resource from which the present resource is derived. The present resource may be derived from the Source resource in whole or part.)	
Location (Spatial Extent)	Geographic Description (Geographic location of the dataset by four coordinates, or geographic identifier)	**COVERAGE** (The extent or scope of the content of the resource. Coverage will typically include spatial location, geographic coordinates, place name or jurisdiction such as named administrative entity)	**Spatial** (The range of spatial applicability of a dataset, e.g., for a dataset of New York weather, the state of New York.)
			SpatialCoverage (indicates areas that the dataset describes: a dataset of New York weather would have spatialCoverage which was the place: the state of New York.)
			Geo (The geo coordinates of the place)
Geographic Feature Type	Spatial Representation Type (Method used to spatially represent geographic information)	**TYPE** (The nature or genre of the content of the resource. Type includes terms describing general categories, functions, or aggregation levels for content)	**Encoding** (A media object that encodes this CreativeWork)
Data Model	Resource Format (Provides a description of the format of the resource(s))	**FORMAT** (The physical or digital manifestation of the resource. Typically, Format may include the media-type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource)	**MediaObject** (A media object, such as an image, video, or audio object embedded in a web page or a downloadable dataset i.e., DataDownload.)
			FileFormat (Media type, typically MIME format (see IANA site) of the content e.g., application/zip of a Software Application binary.)
			EncodingFormat (mp3, mpeg4, etc.)

Appendix B. Users' Search Query Terms as Recorded by the Webmaster Tools Application of Bing.

wgs nairobi kenya shapefile
uganda administrative boundaries shapefile
geological map of the sudan
www.lefenya.co.za
eugene rurangwa fao
political map of namibia site:.za
gemstone geology map in sudan
kenya admin boundaries
rwanda map of roads
sudan geological map
utilization of land title in tanzania -pdf
tanzania infrastructures shapefiles
namibia administrative data shapefile
++serena.coetzee@up.ac.za
+serena.coetzee@up.ac.za
#NAME?
shapefile africa countries wgs
hydrology maps sudan
spatial planning tanzania
biggest inland water bodies in africa
ghana map site:.za
hydrological map of sudan wikipedia
vector street maps site:.za
lebenya
africas area and population
administrative bounderies in kenya
congo dr site:.za
field crop production in botswana
shapefile rivers africa
uganda towns shapefile
yafesi wp-/wp-
rail roads of the world
geology of uganda
r fao vail blue ley
water bodies shapefile
un dataset shapefile cities
bountries oferitrea
boundaries of rwanda
somali map site:.za
mil-v-89039
uganda administrative boundaries
rwanda bounding box
bounding box africa
shapefile of africa rivers
water distribution -china "hotmail.com"
drc map site:.za
megadatauganda
list of rivers in dr congo
map of nigeria and agricultural produce
what east african nation has the most inland bodies of water?
hydrological map site:.za
map of africa eritrea site:.za
nfpi for africover
"south sudan" shapefile
aaglmru
african figures site:.za
geological survey maps online site:.za
tekleyohannes wp-/wp-
what is cropping patterns for africa
agricultural product in nigeria site:.za

hydrological maps site:.za
"http://theacrc.info" site:za
geology south west uganda
africa water bodies shapefile
congo dr map site:.za
administrative boundaries kenya
rwanda administrative map
property boundary maps online site:.za
ghana grographical extent
namibia adminstrative boundaries
riversofafrica
rivers of tanzania
dr congo map site:.za
www.sudanmc.com
kenya towns
cropping patterns in east africa
africa inland lakes shapefiles
rivers of rwanda
geological maps online site:.za
hydrology south africa site:.za
hydrological data site:.za
namibia shapefile mme
namibia shapefile
shapefile african rivers
statial data site:.za
tanzania leben
map of nigeria showing distributions of common cultivated crops
tanzania road map site:.za
rwanda administrative boundaries
yafesi okia linkedin
+land land resources use site:.za

References

1. Nebert, D. (Ed.) Developing Spatial Data Infrastructures: The SDI Cookbook, Version 2.0, 2004. Available online: http://gsdiassociation.org/images/publications/cookbooks/SDI_Cookbook_GSDI_2004_ver2.pdf (accessed on 1 June 2017).
2. Lopez-Pellicer, J.F.; Béjar, R.; Zarazaga-Soria, F. Providing Semantic Links to the Invisible Geospatial Web. In *Notes in Geoinformatics Research*, 1st ed.; Prensas Universitarias de Zaragoza: Zaragoza, Spain, 2012.
3. Vockner, B.; Mittlböck, M. Geo-Enrichment and Semantic Enhancement of Metadata Sets to Augment Discovery in Geoportals. *ISPRS Int. J. Geo-Inf.* **2014**, *3*, 345–367. [CrossRef]
4. Nebert, D.; Whiteside, A.; Vretanos, P.P. (Eds.) OpenGIS® Catalogue Services Specification. Open Geospatial Consortium (OGC), 2007. Available online: http://www.opengeospatial.org/standards/cat (accessed on 5 May 2017).
5. Lopez-Pellicer, J.F.; Florczyk, J.A.; Nogueras-Iso, J.; Muro-Medrano, P.; Zarazaga, J.F. Exposing CSW catalogues as Linked Data. In *Geospatial Thinking, Lecture Notes in Geoinformation and Cartography (LNG&C)*; Painho, M., Santos, M.Y., Pundt, H., Eds.; Springer: Berlin, Germany, 2010; pp. 183–200.
6. Hou, D.; Chen, J.; Wu, H. Discovering Land Cover Web Map Services from the Deep Web with JavaScript Invocation Rules. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 105. [CrossRef]
7. Huang, C.; Chang, H. GeoWeb Crawler: An Extensible and Scalable Web Crawling Framework for Discovering Geospatial Web Resources. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 136. [CrossRef]
8. Purcell, K. Search and Email Still Top the List of Most Popular Online Activities: Two Activities Nearly Universal among Adult Internet Users, Pew Research Center's Internet & American Life Project, 2011. Available online: http://www.pewinternet.org/2011/08/09/search-and-email-still-top-the-list-of-most-popular-online-activities/ (accessed on 4 May 2017).
9. McGee, M. By the Numbers: Twitter vs. Facebook vs. Google Buzz, Search Engine Land, 2010. Available online: http://searchengineland.com/by-the-numbers-twitter-vs-facebook-vs-google-buzz-36709 (accessed on 1 March 2015).
10. Cahill, K.; Chalut, R. Optimal Results: What Libraries Need to know about Google and Search Engine Optimization. *Ref. Libr.* **2009**, *50*, 234–247. [CrossRef]
11. Ochoa, E. An Analysis of the Application of Selected Search Engines Optimization (SEO) Techniques and their Effectiveness on Google's Search Ranking Algorithm. Master's Thesis, California State University, Long Beach, CA, USA, May 2012.
12. Swati, P.; Pawar, B.V.; Patil, A.S. Search Engine Optimization: A study. *Res. J. Comput. Inf. Technol. Sci.* **2013**, *1*, 10–13.
13. Onaifo, D.; Rasmussen, D. Increasing libraries' content findability on the web with search engine optimization. *Libr. Hi Tech* **2013**, *31*, 87–108. [CrossRef]
14. Mustafa, R.; Naawaz, M.S.; Lali, M.I. Search engine optimization techniques to get high score in SERP's using recommended guidelines. *Sci. Int.* **2015**, *27*, 5079–5086.
15. Meta Tags that Google Understands. Available online: https://support.google.com/webmasters/answer/79812?hl=en (accessed on 22 December 2014).
16. Malaga, R.A. Search Engine Optimization-Black and White Hat Approaches. *Adv. Comput.* **2010**, *78*, 1–39.
17. Zhang, J.; Dimitroff, A. The impact of webpage content characteristics on webpage visibility in search engine results (Part I). *Inf. Process. Manag.* **2005**, *41*, 665–690. [CrossRef]
18. Zhang, J.; Dimitroff, A. The impact of metadata implementation on webpage visibility in search engine results (Part II). *Inf. Process. Manag.* **2005**, *41*, 691–715. [CrossRef]
19. Krutil, J.; Kudelka, M.; Snasel, V. Web page classification based on Schema.org collection. In Proceedings of the Fourth International Conference on Computational Aspects of Social Networks (CASoN), Sao Carlos, Brazil, 21–23 November 2012.
20. Dublin Core Metadata Initiative. Available online: http://dublincore.org/ (accessed on 15 November 2013).
21. Harper, C.A. Dublin Core Metadata Initiative: Beyond the element set. *Inf. Stand. Q. Winter* **2010**, *22*, 1. Available online: http://www.niso.org/publications/isq/free/FE_DCMI_Harper_isqv22no1.pdf (accessed on 1 June 2017). [CrossRef]
22. Beall, J. The Death of Metadata. *Ser. Libr.* **2008**, *51*, 55–74. [CrossRef]

23. Google Search Engine Optimization Starter Guide. Available online: http://static.googleusercontent.com/media/www.google.com/en//webmasters/docs/search-engine-optimization-starter-guide.pdf (accessed on 1 November 2014).
24. Zhang, J.; Dimitroff, A. Internet search engines' reponse to metadata Dublin Core implementation. *J. Inf. Sci.* **2004**, *30*, 310–320. [CrossRef]
25. Weibel, S.L. An Introduction to Metadata for Geographic Information. In *World Spatial Metadata Standards*, 1st ed.; Moellering, H., Aalders, H.J.G.L., Crane, A., Eds.; Elsevier Ltd.: Oxford, UK, 2005; pp. 493–513.
26. Katumba, S.; Coetzee, S. Enhancing the online discovery of geospatial data through taxonomy, folksonomy and semantic annotations. *S. Afr. J. Geomat.* **2015**, *4*, 339–350. [CrossRef]
27. Nogueras-Iso, J.; Zarazaga-Soria, F.J.; Lacasta, J.; Bejar, R.B.; Muro-Medrano, P.R. Metadata standard interoperability: Application in the geo-graphic information domain. *Comput. Environ. Urban Syst.* **2004**, *28*, 611–634. [CrossRef]
28. Ali, R.; Sufyan Beg, M.M. An overview of web search evaluation methods. *Comput. Electr. Eng.* **2011**, *37*, 835–848. [CrossRef]
29. Croft, B.; Metzler, D.; Strohman, T. *Search Engines: Information Retrieval in Practice*, 1st ed.; Addison-Wesley Publishing Company: Boston, MA, USA, 2009; pp. 308–338.
30. Can, F.; Nuray, R.; Sevdik, A.B. Automatic performance evaluation of web search engines. *Inf. Process. Manag.* **2004**, *40*, 495–514. [CrossRef]