Master Thesis

RDFa as Semantic Markup and Web Visibility

By
Muhammad Naeem
Omar Tariq Dalal Bashi

Supervisor:
Marie Gustafsson Friberger
Abstract

Web visibility is the appearance of web sites in search engines. Web visibility in search engine is an important factor to improve the e-commerce on the web. If the web site gets high ranking in search engines it will attract more web traffic. Semantic markup is a technique to structure a web site, so it can be understandable by humans and computers. This allows the crawler or spider to understand the content of the web site during the search engine process. Semantically structured web sites increase the web visibility in search engines. RDFa is a semantic markup and supported by the W3C.

In this thesis we have focused on the RDFa as a semantic markup technique. This study shows two aspects of RDFa i.e. what are the benefits and barriers of using RDFa in structuring and enhancing the web visibility of web sites in search engines, and how web developers implement RDFa. This study is based on the data that has been collected through literature review and interviewing different web developers from different companies. First result of this study shows the benefits and barriers of using RDFa according to the web developers. Second result is a guideline for helping the companies that are planning to implement the RDFa in structuring their web sites. The guideline is based on the technical steps and the requirements for implementing RDFa that web developers have described during the interviews.

Key words: Web visibility, Semantic markup, SEO process, RDFa.
Contents

Chapter 1. Introduction ................................................................. 5
  1.1 Introduction and Background ................................................ 5
  1.2 Motivations ........................................................................... 6
  1.3 Goals .................................................................................. 7
  1.4 Research Questions ............................................................... 7
  1.5 Expected Results ................................................................. 7
  1.6 Delimitations ....................................................................... 7
  1.7 Outline ................................................................................. 8
Chapter 2. Literature Review ......................................................... 9
  2.1 E-business and E-commerce ................................................ 9
  2.2 Web Visibility ....................................................................... 10
    2.2.1 The SEO and Search Engine Process ............................... 11
    2.2.2 Web Site Visibility Evaluation ....................................... 13
    2.2.3 Web Visibility and Web Site Structure ......................... 14
  2.3 Semantic Web Technology .................................................... 14
    2.3.1 Ontology ........................................................................ 15
    2.3.2 RDF ............................................................................. 16
    2.3.3 RDFS .......................................................................... 18
    2.3.4 Semantic Markup .......................................................... 18
    2.3.5 RDFa ........................................................................... 19
    2.3.6 Benefits and Barriers of RDFa ........................................ 21
    2.3.7 Samples of Ontologies that can be implemented in XHTML by using RDFa: .......................................................... 23
  2.4 Semantic Web Technology for Web Visibility ....................... 25
    2.4.1 GoodRelations and SEO .............................................. 25
    2.4.2 Use of Semantic Markup by Search Engines ................ 26
  2.5 Summary of literature Review ................................................. 28
Chapter 3. Research Methods ....................................................... 29
  3.1 Method Selection ................................................................. 29
  3.2 Interviews ............................................................................. 30
    3.2.1 Interview Structure and Guide ...................................... 30
3.2.2 Conducting the Interviews ................................................................. 31
3.3 Data Analysis and Presentation .......................................................... 32
3.4 Research Quality .................................................................................. 32
  3.4.1 Validity and Reliability ................................................................... 33
  3.4.2 Ethics .............................................................................................. 33
Chapter 4. Results .................................................................................. 35
  4.1 Interview Discussions ........................................................................ 35
    4.1.1 Interviewees .................................................................................. 35
    4.1.2 Motivation to use RDFa ............................................................... 35
    4.1.3 Requirements to implement the RDFa ......................................... 36
    4.1.4 Technical steps to implement the RDFa ....................................... 36
    4.1.5 External Vocabularies ................................................................. 37
    4.1.6 Barriers of RDFa ......................................................................... 37
    4.1.7 Semantic markup and web visibility ......................................... 37
    4.1.8 Evaluation of web visibility ....................................................... 38
    4.1.9 Benefits of RDFa ....................................................................... 38
    4.1.10 Difficulties of RDFa ................................................................. 38
  4.2 Interviews Results ............................................................................... 41
    4.2.1 Benefits and Barriers of Using RDFa .......................................... 41
    4.2.2 Guideline ..................................................................................... 42
  4.3 Summary of Results ........................................................................... 49
  4.4 Conclusion ......................................................................................... 50
  4.5 Threats to Validity ............................................................................. 50
Chapter 5. Discussion ............................................................................. 51
Chapter 6. Conclusion and Future Work .................................................. 54
References .............................................................................................. 55
Appendix 1: Interview Guide .................................................................... 59
Appendix 2: Abbreviations and their Definitions ..................................... 61
Appendix 3: XHTML ............................................................................... 62
Appendix 4: Interviews’ Summaries ......................................................... 64

Figures
  Figure 1: Thesis Outline ........................................................................ 8
  Figure 2: The Crawling Process [19] ..................................................... 11
  Figure 3: The Search Engine Process ................................................... 12
Figure 4: An example of an ontology, consisting of classes and subclass relationships ................................................................. 16
Figure 5: Giving meaning to the nesting tags [11] ................................................. 16
Figure 6: RDF example [11] ........................................................................ 17
Figure 7: RDF class and property being described by RDFS [11] ............... 18
Figure 8: Compact the URIs to a prefix [29] .................................................. 20
Figure 9: Using the compact URIs [29] ....................................................... 20
Figure 10: The GR ontology is used to describe the restaurant web site and name [12] ........................................................................................................ 21
Figure 11: How FOAF describes a person [13] ............................................ 23
Figure 12: how objects can be described in DC [10] ................................. 24
Figure 13: Converted URI to CURIE .......................................................... 45
Figure 14: Usage of GoodRelations for adding company information .......... 45
Figure 15: Usage of FOAF for adding personal information ...................... 46
Figure 16: Flowchart of the guideline to implement RDFa ....................... 48

Tables
Table 1: The interview’s findings ................................................................ 40
Table 2: Benefits and barriers of using RDFa ........................................... 41
Table 3: Benefits of RDFa with respect to each interviewee .................... 42
Table 4: Barriers of RDFa with respect to each interviewee ...................... 42
Chapter 1. Introduction

1.1 Introduction and Background

It can be a primary objective of any web sites owner is to increase web visibility of their web sites in search engines. Most searchers use the results that appear on the first page in search engines without going further to the second or third results page [23][7]. We are considering the web visibility in online visibility meanings “Online visibility can be defined as the extent to which a user would come across an online reference to a company’s web site” [32]. Web visibility can be measured by the performance of the web site in search engines i.e. the position of the web site in the search engine’s results. Web visibility in search engines is an important factor in improving the e-business on the web because if the web site gets a high ranking in search engines then it will attract more web traffic to the website [23][18][32].

There are different methods to increase the visibility of web sites in search engines, and the structure of the web site is an important component in making the web site's code more machines readable [23][1]. Semantic markup is a way to structure a web site so that it can be understandable by humans and computers [23]. In semantic markup we use special tags to designate the sections and content of the web pages. This allows a crawler or spider (a software program that is used during the Search Engine process) to understand the content of the web during the search engine process. In this way we increase the web visibility of our web pages in the search engines. By using the semantic markup the search engines can know the contents of the web pages. With the help of semantic markup search engines determine the topic and relevancy of different sections of the web pages [23].

There are different semantic markup techniques like microformats and RDFa. Microformats are semantic markup technique and they are composed of simple set of data formats. Microformats are built on the existing standards and they are used to solve simple problems. RDF triples and external vocabularies can not be used with microformats because microformats are using their own predefined rules. There are separate parsing rules for each of microformats and they can not integrate in XML language [1].
Resource Description Framework in attributes (RDFa) is a semantic markup that can communicate with the crawler of a search engine [22]. When we use RDFa rich snippets structure data in our website then we gain more control over the way in which our website will appear in the search engine [22]. RDFa provides a way for the Extensible Hyper Text Language (XHTML) authors to design human readable data that can be interpreted by the browsers and other programs. RDFa is specified for XHTML 1.1 because RDFa is extensible. RDFa benefits from the RDF that is W3C standard for interoperable machine readable data. There are different attributes in XHTML that are relevant to RDFa [3]. Consider sections 2.3.2 and 2.3.5 for further explanation for RDF and RDFa.

It has been observed that many companies pay huge efforts and attention to enhance their website’s visibility on the web. By using semantic markup the website structure becomes machine readable, as a result of which website’s contents are understandable by the crawler or spider [14]. We believe that by showing the benefits and barriers of semantic markup, and by creating a guideline about how semantic markup can be implemented, many companies may encourage implementing this technology in their websites.

1.2 Motivations

The processes for websites to get high rank or making it more visible among the top results in major search engines were not hard in early days of search engine optimization. The search engines algorithms were easy to follow. It was so easy that a developer just need to include the keyword he/she wanted to rank in the title tag of webpage, and spread this keyword all over the page content. Now a days, search engine algorithms are more complex [6]. It has been proposed that semantic markup can play a big role in raising website ranking and enhancing website visibility in search engines [16]. Many companies emphasis on the style of their websites not on the structure of the code, it is good for human that a website is looking good but it is not understandable by the browsers. The crawler or spider in search engine process may not understand the content of the website clearly; as a result websites get poor visibility in search engines [14]. In order to structure the website semantically, there are different semantic markup techniques. RDFa is a relatively new semantic markup technique, so there is a need to explore the knowledge about RDFa [16].
1.3 Goals

The first goal of this thesis is to determine the benefits and barriers of RDFa in structuring the web sites and enhancing their visibility in search engines. The second goal is to create a guideline that helps companies to use RDFa as semantic markup for structuring the code of their websites to enhance companies’ websites visibility in search engines. These goals will be achieved by interviewing web developers from different companies that are using RDFa in structuring their web sites.

1.4 Research Questions

We formulate two research questions as follows:

- What are the benefits and barriers of using RDFa in structuring web sites and enhancing their web visibility?
- How do web developers implement RDFa in structuring web sites?

1.5 Expected Results

We expect this thesis provide knowledge about semantic markup in particularly RDFa with its benefits and barriers, and its role in enhancing the web visibility. The other expected outcome will be a guideline for the companies that want to use semantic markup in structuring their web sites. This guideline will show how companies can implement RDFa as a semantic markup technique. The results will be based on the experiences of web developers who are using RDFa as a semantic markup.

1.6 Delimitations

There are different semantic markup technologies in the market like microformats and RDFa. We limit our study to find how developers implement RDFa and what the benefits
and barriers of RDFa are. We choose to study RDFa since external vocabularies like GoodRelations, Dublin core and FOAF can be used with RDFa but it can not used with microformats. RDFa is a W3C recommendation as compared to microformats [16].

1.7 Outline

Chapter 2 presents a literature review of Semantic Web and Web Visibility. Chapter 3 describes the Research Methods that have been used in our qualitative study. Chapter 4 presents results of the study, which includes interview discussions, interview results, summary of results, threats to validity and conclusions. Chapter 5 presents a discussion of our thesis results. Finally, chapter 6 presents the conclusion and the future work of our study.

![Figure 1: Thesis Outline](image)
Chapter 2. Literature Review

In this chapter, firstly, basic concepts related to e-business are presented. Then, Web Visibility and SEO are introduced, and finally Semantic Web and Semantic Markup are discussed. The fact that much information about our main topic of the thesis is contained in online blogs is a problem for the topic and our work. Our study is needed to provide more rigorous exploration of the topic. In the literature review chapter we discuss authorized articles and books related to our topic.

2.1 E-business and E-commerce

E-business is application of telecommunication and information technology that work together for conducting businesses [26]. E-business involves connecting partners, suppliers, providers and consumers by using the internet. It aims to use the same business strategies as in the real market and improve these strategies to be more efficient [26].

E-commerce is narrower than e-business, and focuses only on the buying and selling of products and services on the internet. There are many types of e-commerce such as business between enterprises (B2B), business between enterprises and consumers (B2C) and business between consumers performed through the (C2C). Since many e-business processes are performed through the companies’ web sites, these are the interfaces and the main gate to any e-business platform [26].

For this reason, the web visibility of a web site in search engines is very important in the e-business area. If a company lacks web visibility, then potential customers might not find it [18]. The following section describes aspects of web visibility.
2.2 Web Visibility

Web visibility is the extent of the web site to be seen by users [32]. Mostly, the first results in search engines get more traffic from users which lead to more benefits to e-business web sites owners. Since web search engines are a main source of information for most web users, they are increasingly important for e-business [18].

To be the first result in search engine may be the goal of any web site owner. In order to reach this goal, a web site owner has to make his/her web site more visible on the web. Most searchers make use of the results that appear on the first results page in search engines without going further to the second or third results page [7]. Another factor that affects web visibility is giving a meaningful appearance to the results. Just to be in the first page of results is not enough to let the web site get more clicks. The web site has to get the searchers attention by working on giving the results of web site some meaningful appearance like a photo or an effective title [7].

Web visibility can be measured by the ranking of a web page in search engines. Web visibility may be an important factor in e-business, if the web site is not visible in search engines then business lose a lot of customers as we mentioned in 2.1. In order to increase web site visibility, the first action to be taken is to evaluate the current positioning of the web site in the search engines. For this purpose web analytics software is used [18] [32].

Some technique has to be added to the web site to make it visible enough among millions of web sites on the net. Especially, many web sites have many matched information, design and content [18]. In our thesis we will focus on how this technique can be added to the web site’s structure.

The following sections describe how web structure influences the web visibility, how to evaluate the web visibility and the search engine process.
2.2.1 The SEO and Search Engine Process

Search engine optimization (SEO) is the process of improving the visibility of a web site or a web page in search engines. SEO is an art of driving web traffic to the web site without paying for each click that comes to the web site through the search engines [1].

To be able to implement SEO on web site, the developer has to have some knowledge about search engines and their work mechanisms. Google search engine and most other search engines use the same mechanism [14]. We take the Google search engine as an example to explain how search engines work. The following steps explain the google search engine process and Figure 3 illustrates the process.

**Step 1:** Discover and find web sites by crawling to the web sites through their links. In this step Google uses a software program called crawler also known as spider. A crawler is used to browse the web pages in an orderly fashion and this process is known as crawling. The crawler finds the web pages and the other links in the same page through crawling. It starts with a seed web page and from this seed page starts crawling by using the links found in the seed page [14] see Figure 2.

![Figure 2: The Crawling Process](Image)

**Step 2:** Google search engine stores the keywords, the summaries and the information of the web pages through the Index server system, also called the Indexer. In this step the Google search engine has the exact copy of every page that the crawler has found [14]. The Indexer takes the full pages’ text found by the Crawler and saves it in Google’s index database, which stores the index alphabetically with its location and where it appears.
Structuring data as mentioned in step 2 lets Google’s search engine find the query requested text faster [13].

**Step 3**: Ranking the web pages that are found in Index. The algorithm which Google uses for ranking the web pages is known as PageRank. PageRank is based on the idea that has been used by librarians in the pre-Web past to score articles and other scholarly documents. If one document has more citations to other documents it makes the document more important with high rank. In the PageRank algorithm, the ranking of the web page depends on the number of web pages that are linked to this page. There are two kinds of links: inbound and outbound links. As an illustration, a link from web site A to web site B is the outbound link from A and an inbound link to B. In PageRank, a web page has higher rank if it has higher inbound links. In the other words, ranking of a web page is directly proportional to inbound links of that page [14].

**Step 4**: The search engine returns the results of web pages to a specific query from the index. Web pages in the index are ranked according to the PageRank algorithm [14].

![Diagram](http://example.com/search-engine-process.png)

**Figure 3**: The Search Engine Process

**SEO Methods**

There are mainly two kinds of methods in SEO i.e. white hat SEO and black hat SEO. These methods can be used in order to optimize the web site visibility in search engines. The web sites that are using the black hat SEO method are at risk to be banned by most
search engines because it is not a recommended method. Most search engines created different software to avoid the black hat SEO methods that aim to trick the search engines. The recommended SEO method is white hat SEO [19]. White hat and black hat SEO methods are explained as follows [19]:

White hat SEO

- Make the web site include meta tags, photos, information and key words that help the search engine to understand what the web site content is about.
- Describe and include all the relevant links that refer to web site from other web sites.
- Include advertising key words in the web site like marketing offers, these key words work on increasing the traffic to the web site.
- Submit the web site to the search engines manually without waiting the crawler to crawl it.

Black hat SEO

- Embed hidden key words in the web site in order to trick the crawler.
- Create inbound links from other unrelated web sites to higher ranking in search engines.
- Submit the web site repeatedly to the search engine in order to get in order to get higher ranking.

2.2.2 Web Site Visibility Evaluation

Web site visibility evaluation is an important step in enhancing web site visibility in search engines. Through the web site evaluation we will know the position of the web site in the results of search engines (is it visible enough or not). There are many tools and methods (like Google Analytics1, AWstats2 and eLogic3) that can be used to analyze web site

---

1 Google Analytics: https://www.google.com/accounts/ServiceLogin?service=analytics&userexp=signup&hl=en
2 Awstats: http://awstats.sourceforge.net/docs/index.html
3 eLogic: http://www.elogicwebsolutions.com/
visibility. These tools analyze the web site visibility by focusing on three aspects: the number of the visitors to the web site, the links to the web site from other sites and how the web site performs in search engines. These three aspects can be used to analyze web site visibility, which helps to extract the drawbacks in the web site and which area needs to be improved to make the web site more visible [18].

2.2.3 Web Visibility and Web Site Structure

The tremendous growth of web technology has increased the amount of information available on the internet. A search engine may do a good job in indexing the web pages but in most cases the search engine software cannot read and understand the exact meaning of web page content [16]. To explain this problem we take an example, if we suppose to query a search engine using the three terms “book”, “about” and “hotel”. From this query, it is clear for a human reader that we want a book about hotels, but search engine displays results related to hotel booking. To solve this problem semantic markup (which will be further presented in section 2.3.4) is a possible solution because semantic markup makes the web page code more machines readable. The web sites that are structured semantically get higher ranking in the search engines because the contents of these web sites are more understandable to the crawler during the SEO process. By providing semantic structure to a web site, its visibility in search engines may be increased [16].

2.3 Semantic Web Technology

The World Wide Web has changed the way of communication among the people and the way of conducting businesses. The present web’s contents represent the information to be more human readable and understandable rather than machine readable. The semantic web is the web of data rather than the web of documents. Semantic web is machine readable [34]. Adding semantics to web site structure makes the web site code readable by both humans and machines. The semantic web contains meta-data, which is data about data and it contains ontologies. Ontology is an agreement needed to be added to the web page to let the machine understands the document [16]. The Resource Discretion Framework (RDF) gives users the opportunity to describe the resources by their own ontology by using the Resource Discretion Framework Schema (RDFS) language. RDFS is responsible for
defining the vocabulary of domain [11]. By using meta-data and ontologies, semantic technology adds meaning to the web page. The benefit of semantic markup can be noticed in the search engines results [16].

The following sections explain the technologies related to semantic web. These technologies are ontologies, RDF, RDFs, semantic markup RDFa etc.

2.3.1 Ontology

The term ontology originates from philosophy as “the study of the nature of existence” [11], which is about describing the things that exist in the world around us. In computer science, ontology has a different definition: “an explicit and formal specification of a conceptualization” [35]. Ontology includes classing and subclassing (see Figure 4) of describing a domain and its concepts with their properties and the interrelationships between these concepts by including information about the domain such as [11]:

- Classes of objects of the domain (movies, actors, directors)
- Relationships between these classes or the class hierarchy (X is an actor in Y)
- Properties (for example movie X is produced by Y)
- Value restrictions (for example only directors can direct a movie)
- Disjointness statements (for example movies and actors are disjoints)
- Logical relationships between domain’s classes (for example each movie must include at least one director).
2.3.2 RDF

RDF is a graph framework to represent the information and to give metadata about the resources on the World Wide Web. It is a data model that consists of triples. Triples are the RDF statements. RDF statement consists of object, attributes and value and that is why RDF’s statement is called a triple [11].

The RDF language was created to give metadata about the resources on the web. The need for RDF increased because of XML’s drawbacks in giving meaning to the data in the nested tags. There is no standard tactic to explain this nesting but each application uses its own tactic. Figure 5 shows how meaning can be added to the nesting tags; Jhon Black is a lecturer of English literature.

```xml
<course name="English Literature">
    <lecturer>John Black</lecturer>
</course>
<lecturer name="John Black">
    <teaches>English Language</teaches>
</lecturer>
```

**Figure 5:** Giving meaning to the nesting tags [11]

**Basic ideas of RDF:**

The basic idea of RDF is the object- attribute- value triple. Such a triple is called a statement. In other words, it is a resource, property and value, which can be either resources or strings. A triple consists of resources which may be an object as a car or shop name.
Each one of those resources has a Universal Resource Identifier (URI), such as a Uniform Resource Locator (URL) or any other identifier. Furthermore RDF has properties. Properties are a type of resources which are also identified by a URI. These properties are used to describe the relations between the resources like owned by, color, name, etc [11]. RDF uses the syntax of XML, such as all the attributes must be written in small letters, all values must be between quotations, etc [11].

There are three ways to view the RDF’s statement [11]:

1- Triple or set of triples. For example: if we take the triple (David, P, Jenny), it is the same as P (David, Jenny). The predicate P represents the relation between the objects David and Jenny. RDF can only relate two objects by a binary predicate. For example the binary predicate P= David Billington relates the two objects X= http://www.cit.gu.edu.au/~db and Y= http://www.mydomain.org/site-owner in this way X, Y, #P.


2- Graphical representation: it is the way to represent the triple by drawing labeled nodes that are connected by arcs, those nodes represent the subject (the resource) and the object (the value) in the triple, and the arcs represent the predicates between these nodes. For example:

www.cit.gu.edu.au/~db______ site-owner______David Billington

3- XML code or document: this type of statement representation is based on XML but XML is not included in the RDF data model. See Figure 6 for an example:

```xml
<rdf:RDF
 xmlns:rdf=http://www.w3.or/1999/02/22-rdf-syntax-ns#
 xmlns:mydomain=http://www.mydomain.org/my-rdf-ns>
 <rdf:description
 rdf:about="http://www.cit.gu.edu.au/~db">
  <mydomain:site-owner rdf:resource="#David Kage"/>
 </rdf:Description>
</rdf:RDF>
```

Figure 6: RDF example [11]
2.3.3 RDFS

RDF Schema is a language used for describing semantically the classes and the properties of RDF domains. Furthermore, RDFS gives developers the ability to define his/her own RDF ontology, the properties of each object, the relationships between objects and the optimal value which may take each object [15]. Figure 7 gives an example of a University (courses and lecturers) in RDFS.

```xml
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <rdfs:Class rdf:ID="lecturer">
    <rdfs:comment>
The class of lecturers
All lecturers are academic staff members.
    </rdfs:comment>
    <rdfs:subClassOf rdf:resource="#academicStaffMember"/>
  </rdfs:Class>

  <rdfs:Property rdf:ID="involves">
    <rdfs:comment>
It relates only courses to lecturers.
    </rdfs:comment>
    <rdfs:domain rdf:resource="#course"/>
    <rdfs:range rdf:resource="#lecturer"/>
  </rdfs:Property>
</rdf:RDF>
```

Figure 7: RDF class and property being described by RDFS [11]

2.3.4 Semantic Markup

Semantic markup is a technique to structure the web site semantically, so the web sites are understandable by human and computers [1]. There are different techniques that are used to structure web sites semantically. The most popular techniques are Microformats and RDFa. Microformats and RDFa share the same goals but they are quite different from each other in the aspect of implementation [1]. In our research we focus on RDFa, because it’s the preferred technique of the W3C and because it’s more stable and powerful than microformats [16]. RDFa derives its power from the ontologies that it is based on, like Friend of A Friend, GoodRelations and Dublin Core [1]. See 2.3.7, samples of ontologies.
A semantic markup may include an RDF document that contains RDF statements and may contain many different vocabularies. When the semantic markup document is added to a web page, it describes the content of the web page with the help of the defined keywords in the vocabularies used in the RDF file. Whenever the crawler reaches a web page that contains a markup file, it loads the markup file with the included vocabularies. At this stage the crawler or any other application behaves as it understands the web page content and it discovers the important keywords are predefined in the RDF statement. As a result the web page is not only human readable but machine understandable also [23].

Section 2.3.5 explains RDFa. Before explaining RDFa we have to understand some technologies that are related to RDFa, markup technologies like RDF, ontologies and RDFS. We present these different technologies in order to give the reader a full image about RDFa.

2.3.5 RDFa

Semantic web is a web of data more than a web of documents. For this propose, we need to structure and design our web sites to be machine and human readable. This can be possible by using RDFa. RDFa stands for resources descriptions framework in attributes and it was developed by W3C. RDFa can add semantic information to the XHTML markup by reusing the attributes that are already available in XHTML and apply them to the other parts. By using RDFa, RDF triples can be embedded in the XHTML document which gives the ability to embed several vocabularies in the XHTML document. Furthermore, it could be easier for the web developers to extract the RDF triples from a web page that is structured with RDFa [1]. RDFa gives the ability to embed the structured data in XHTML. RDFa is a markup that reuses the rendered and hypertext data of XHTML, so the developers do not repeat themselves [30].

RDFa makes use of RDF triples which are self contained in RDFa. Self Containment makes the RDF triples decoupled from the XHTML code [4]. RDFa has some specific attributes as [5]:

- The @about attribute is used to represent the subject.
- The @property attribute is used to represent the value.
• The @resource attribute is used to represent the object.
• The @datatype attribute is used to represent the datatype of the resource.
• The @typeof attribute is used to represent the type of the resource.

URIs use to identify the location of any XHTML document when it is being published on the web. RDF deals with a full URIs (not relative paths), so it is not possible to use URIs for representing the RDF triples. As long as RDFa is a way for RDF to be embedded in XHTML, then every relative path in the converting of RDFa to triples must be resolved to its origin URIs. Therefore, in RDFa, CURIEs are used. CURIEs is abbreviation for Compact URIs [5].

In CURIEs the leading part of the URIs is changed with a token [29] as in the following example. The full URIs of Albert Einstein on Dbpedia:

http://dbpedia.org/resource/Albert_Einstein

This URI compacted by CURIE to a prefix mapping, and the prefix would be linked to some leading token URIs. In RDFa the XML namespaces are used for this mapping [29]. See Figure 8.

```
<div xmlns:db="http://dbpedia.org/">
  ...
</div>
```

**Figure 8**: Compact the URIs to a prefix [29]

After creating the prefix, the developer can use the compact URI [29]. See Figure 9. The CURIE resolves to a full URI, according to the namespace declaration in the page [21].

```
<div xmlns:db="http://dbpedia.org/">
  <div about="[db:resource/Albert_Einstein]">
    ...
  </div>
</div>
```

**Figure 9**: Using the compact URIs [29]
With the help of RDFa, different ontologies can be used in XHTML [5]. One such ontology is GoodRelations which covers e-commerce concepts. This ontology is described further in section 2.3.7, along with other ontologies commonly used with RDFa.

Figure 10 shows part of code as an example about using both RDFa attributes and GoodRelations vocabulary on a restaurant homepage.

![Figure 10](http://www.example.com/xyz/company)

**2.3.6 Benefits and Barriers of RDFa**

Most of technologies have benefits and drawbacks and the need for those technologies varies from a user to another. Factors involved in choosing a specific technology include why the technology is needed and its intended use. As stated in 1.6, RDFa was chosen as the topic semantic markup technology to study in this thesis. Some benefits and barriers of RDFa are discussed below.

**Benefits of RDFa**

1- RDFa makes the code human and machine readable. Semantic markup adds to the website structure by reusing the attributes that already exist in XHTML, use the own attributes of RDFa and make use of the RDF triples [1].
2- The RDF triples are not coupled to the XHTML code because of the Self Containment in RDFa [1]. All RDFa’s fragments contain a full data structure, which gives the ability to copy and paste these fragments and make the RDF triples decoupled from the code [4].

3- Different vocabularies can be implemented in the web site structuring [1]. The developer can implement his/her own vocabulary and many external vocabularies in structuring his/her web site structure like (GR, FOAF and DC) [30].

4- RDFa is easy to implement. RDFa uses the attributes that already exist in XHTML plus its own attributes which makes it easier to extract RDF triples from an RDFa marked document [1].

5- RDFa is invented and supported by the World Wide Web Consortium (W3C) [1]. W3C is the main international organization working on developing the World Wide Web. This privilege gives RDFa ensured long term work and support factors [39].

6- RDFa uses the DRY idea (Don’t Repeat Yourself) [1]. The DRY principle reduces the possibility of introducing inconsistency. Every piece in the code is presented with complete information only once; the aim of avoiding representation of the same piece of the code many times is to avoid the probability of describing the same feature in two different ways [9].

**Barriers of RDFa**

1- The page content must be written with XHTML 1.1 or later versions. According to W3C, it is not possible to implement RDFa in HTML because it is not an extensible language like XHTML [3].

2- It is not possible to use any XHTML cleaning tool to make the page content wellformed, these tools affect the RDF statements in the code structure [1]. Cleaning tools are used for finding and correcting the errors in the XHTML code like Tidy[^4].

[^4]: [http://infohound.net/tidy/](http://infohound.net/tidy/)
3- A special type of URIs, called CURIEs, have to be used with RDFa [1]. RDF describes the resources that contain a complete URIs. In RDFa, the developer must change the complete URIs to a compact URIs (CURIE). In CURIE the leading part of URI being changed with a token [29].

2.3.7 Samples of Ontologies that can be implemented in XHTML by using RDFa:

Friend of a Friend (FOAF)⁵

FOAF is a vocabulary of persons and their relations. The aim of creating the FOAF ontology was to connect the information that is published by people on the web, specially the documents that contain “see also”, in other words documents that contain links that refer to other documents, which helps the machine to make use of that information and it gives the computer programs the ability to move through a machine readable web [13]. An example of how FOAF can be used is given in Figure 11.

```
<foaf: person rdf:about="$me" xmlns:foaf="http://xmlns.com/foaf/0.1/">  
  <foaf:name>Dan Brickley</foaf:name>
  <foaf:mbox_sha1sum>241021f5b0e6289f92815fc210f9e9137262c252e</foaf:mbox_sha1sum>
  <foaf:homepage rdf:resource="http://danbri.org/"/>
  <foaf:img rdf:resource="http://rdfweb.org/people/danbri/mugshot/danbri-small.jpeg"/>
</foaf:person>
```

**Figure 11:** How FOAF describes a person [13]

Dublin Core (DC)⁶

DC is metadata about different things; such as network references, locations’ information, companies’ names and contacts numbers etc... . DC adds semantic to these things or recourses. DC working on describing the resource by creating a special class of statement, this statement consists of two parts: elements (nouns) such as a title, subject, type, etc. and

⁵ [http://xmlns.com/foaf/0.1/](http://xmlns.com/foaf/0.1/)

qualifiers (adjectives) [10]. Figure 12 gives an example of how objects can be described with DC.

![RDF code](image)

**Figure 12**: how objects can be described in DC [10]

**GoodRelations**

The GoodRelations ontology covers the e-commerce domain, and is often presented as a means of raising web visibility [25]. It has been used by many huge companies like the large electronics company BestBuy, prominent search engines like Google and Yahoo!, the e-commerce web site Overstock.com, OpenLink Software and the online technology book store O’Reilly [12].

We take GoodRelations as an example of vocabularies that can be used in RDFa (see figure 10). GoodRelations ontology covers the reputation needs for E-commerce [25]. GoodRelations can be used for description of business offerings in a precise way. This ontology can also be used for describing the resources and the relationship between them, the data package that has description of products, prices of products, properties of the products, stores, opening and closing hours and mode of payment etc. All these data can be embedded into the web page, which increases in its role the visibility of the web page in search engines [25].

GoodRelations is a vocabulary that can be used for many purposes and it is investigated in Semantic web application and traditional search engine. GoodRelations is a multi-syntax data format because it can be published in different formats like HTML, RDFS etc. [12].

---

7 http://www.heppnetz.de/projects/goodrelations/
2.4 Semantic Web Technology for Web Visibility

The semantic web is an enhancing technology that can be used on different fields on the web [37]. One of the web fields that can be enhanced by semantic web technology is web visibility. A web site to be visible enough on the web, has to have high performance in the search engines because search engines are the primary way to find new web sites for most of web users. Semantic web enrichs the web site code with metadata and ontology which are the main factors that can help the machine to understand the meaning of the document, which may affect in its turn the results of search engines positively. Thus the users of search engines can find their target more easily since the machine understands the queries of the users and orient them to the right direction. Furthermore, the web page that is described semantically may rank higher in search engines’ pagerank because it would be more readable and understandable by search engines software as we mentioned in 2.3.4 [16].

We present GoodRelations and SEO as an example about the overlap between semantic and web visibility.

2.4.1 GoodRelations and SEO

Traditional SEO as explained in section 2.2.1 gives the opportunity for the online retailer to increase web traffic to their web sites. SEO has several limitations. For example, SEO success depends on how the search engines deal with the web site, which semantic in its role plays a positive role in it. The GoodRelations vocabulary can fill this gap and gives an extra advantage to the retailer. GoodRelations increases the products visibility in the latest generations of search engines and provide detailed information about the products [33].

Improvement for BestBuy

BestBuy is using the GoodRelations in structuring their web site. Best Buy reports getting an increase in the web traffic to their web site [12]."GoodRelations + RDFa has improved the rank of the respective pages in Google tremendously. In fact, if we try the query "BestBuy Ferris Bueller" on Google, then the page comes on rank # 1 ahead of the much
more established pages. This indicates a strong effect of GoodRelations + RDFa on Google's appreciation of a page”, [33].

**Used by Search Engines**

GoodRelations can be used for rendering, ranking and in increasing the web visibility of web pages on different search engines and applications. The rendering of web pages in search engines is improved by adding GoodRelations [12].

The web pages that contain GoodRelations in their code get higher ranking in search engines. GoodRelations increase the visibility of the web page in search engines because it adds the business data to the web page in a machine understandable way [25].

GoodRelations has improved rendering in Yahoo search results. Now Yahoo provides detail information on products if the web pages contain GoodRelations in the results of Yahoo SearchMonkey. In Google with the help of GoodRelations, detail information on products like price info of product is available. GoodRelations increase ranking of the web page because of its higher data specificity. GoodRelations has also increased the visibility of web pages on mobile applications like Mobedo⁸ [12].

**2.4.2 Use of Semantic Markup by Search Engines**

Bröcker and Van Ahee [16] evaluated semantic markup enhancing web visibility in search engines. Using a case study on three different search engines they analyzing how these search engines reacted on the web sites that are emantically structured. They chose Google search engine, as the biggest search engine on the web, Yahoo SearchMonkey, as it supports semantics and Hakia search engine because “Hakia claims to be the only true semantic search engine providing results only on concept match rather than keyword match or popularity ranking” [16]. They focused on the effect of three different semantic technologies (microformats, RDFa and RDF) on the result of the above mentioned search engines. The following results were concluded from this project:

⁸ https://www.mobeedo.com/cockpit/current/
Since Google is the biggest search engine in the market with many webmasters and web designers are working on getting as much traffic as possible from it. Google search engine did not show any reaction against meta data. “On the other hand, Google has a reputation for deploying new features at high rate, so they could be possibly working on semantics in the background. This is however pure speculation, there is no concrete evidence” [16].

The Yahoo! Search engine adopts clear future plan about how to go forward in the semantic web field. They are making use of semantic data in their communications and they started teaching web masters and web designers the benefits of semantic web. Furthermore they give the opportunity to web developers to build their own semantic search engines. According to Bröcker and Van Ahee [16], Yahoo tries to be a lead in this domain. It’s clear that they are processing the meta data in the web pages to introduce more information in the results. On the other hand, there is ambiguity about if Yahoo is using meta data in ranking its results which may lead to many developers not being encouraged to use semantic markup in structuring their web sites [16].

Hakia supports semantic web technology through adopting their own ontology, since few web sites offer semantic through the standards ontologies (like FOAF and DC). Using the standard ontologies may conflict with their own ontology, so Hakia decided to avoid working with RDFa, microformats and external RDF. For this reason the authors of this project didn’t analyze those semantic technologies on Hakia [16].

We would like to mention that at the time when this project was done in 2008, Google search engine wasn’t supporting semantic markup and Yahoo was unclear about using semantic in ranking its results. Now a days, Google and Yahoo search engines support semantic markup technology [12]. This study is to some extent is a related work for ours because they had worked with different search engines reactions against semantic markup. As we have mention above that google and yahoo support the semantics. From this study we get idea about our thesis to study how semantically structured web sites affect the web visibility of web sites in search engines.
2.5 Summary of literature Review

In this chapter we have discussed the main areas that are related to our study. In the first section we have discussed about e-business and e-commerce because our study will impact this area. Second section is about web visibility, web visibility and web sites’ structure, evaluation of web visibility and search engine process. Web visibility is the main part of our study because we will investigate how semantic markup enhances the web visibility in search engine. To conduct our study it is very important for us to understand the background of web visibility.

In the third section we have discussed about the semantic web and related semantic technologies like semantic markup, ontologies, RDF, RDFa etc. In order to understand our study it is important to sound knowledge about all these technologies because semantic markup and especially RDFa is a focal point in our study. The last section is about the semantic web technology for web visibility that describes the usage of semantic technologies in enhancing the web visibility in search engines. We have discussed GoodRelations as an example of using semantic technology in order to understand how semantic technology enhances the web visibility in search engines.

Our thesis is an advanced research in the semantic markup field because we will gather the data from what exactly is happening in the practical field of implementing the semantic markup in structuring web sites. We have investigated how semantic markup enhances the web visibility of web sites in search engines. In our thesis, we are focusing on the semantic markup and its role in enhancing web visibility more than focusing on the search engines. RDFa was chosen as a semantic web technology see 2.3.4. For this purpose we have also investigated the benefits and barriers of using RDFa.
Chapter 3. Research Methods

This chapter presents method choice that we have made in our thesis. It provides the reader an opportunity to know about our approach of study and the reasons behind the selection of a specific research method.

3.1 Method Selection

An important aspect of research is to select an appropriate research method to conduct the research. There are mainly two types of research approaches, quantitative and qualitative. The empirical data produced by these approaches differ from each other [8]. The qualitative research method has its origin in the social sciences. It is concerned with increasing the knowledge and understanding of subject rather than producing explanations for it. Qualitative research methods are common in the area of information sciences and interview is a useful technique to gather the qualitative information [8]. The qualitative research method is an appropriate method for creating understanding and it is a suitable method for dealing with the complex questions since it gives more specific information from a single respondent [17].

In our research we have selected the qualitative research method due to the nature of our research questions. How do web developers implement RDFa in structuring web sites? And: What are the benefits and barriers of using RDFa in structuring web sites and enhancing their web visibility? The research question in qualitative research often start with how and what [17]. The quantitative research method is not appropriate with our problem because of the nature of the research questions and the exploratory nature of the study. Therefore we have selected interviews as a qualitative research method for this study.

We investigated what are the benefits and barriers of using RDFa and how companies can use RDFa as semantic markup. In interviews, we can work directly with the respondent and it is generally easier for respondent, especially if what are sought, are opinions or impressions. Interviews provide the interviewer the opportunity to explore and investigate the topic in-depth through asking follow-up questions. We can conduct telephonic
interviews, internet-based interviews etc., but it depends on the availability of the interviewees.

3.2 Interviews

We selected interviews as a data collection method for our thesis. For our investigation we selected the web developers from companies who are using RDFa as a semantic markup.

3.2.1 Interview Structure and Guide

There are mainly four types of interviews and those are structured, unstructured, semi-structured, and group interviews [20]. In structured interviews, the interviewer has some pre-set questionnaire, interviewer gets more specific answers for his/her questions because for the interviewee it is difficult to move away from the main agenda. In unstructured or open interviews, the interviewee can extend his/her answers without any constraint but he/she can move away from the agenda, so we have to keep in mind to draw him/her back to the main agenda of the interview [8]. Semi-structured interviews are a combination of structure and unstructured interviews. Group interview involves a small group guided by an interviewer who facilitates discussion on a specified set of topics [20].

The most appropriate type of interview depends on the questions to be addressed, the goal of the interview and the research method. If the goal of the interview is to gain an overall understanding of a subject, then unstructured interview is often a suitable approach. But if the goal of the interview is to get knowledge and understanding about a specific issue or topic, then a structured interview is often a better approach [20].

In our thesis, we have specific issues i.e. benefits and barriers of RDFa and how web developers implement the RDFa as semantic markup. The aim of our thesis is to increase the understanding of semantic markup and to create a guideline that will help companies to use RDFa as semantic markup and the benefits and barriers of RDFa. For creating a guideline we need specific answers for specific questions related to our topic. Therefore, we have selected the structured interview for our thesis. We have used the structured interviews for our research and we have developed an interview guide which contains close
ended questions. The interview guide was created in the light of our needs for our thesis’ results. Firstly, the steps of the guideline for implementing RDFa that we would create will be used for helping companies that want to implement RDFa in their web site structure. Secondly, present the benefits and barriers of RDFa, see appendix 1.

3.2.2 Conducting the Interviews

We needed to find web developers that have good experience in RDFa as semantic markup. Our purpose was to investigate their experience in this area by finding out how they implement and work with RDFa. Since semantic markup is a fairly new technology, there are few experienced web developers in semantic markup. We paid a huge effort in trying to recruit some interviewees through contacting a list of companies that may use RDFa in structuring their web sites. Unfortunately, we didn’t get any answer from these companies. Interviewees were recruited through personal contacts within the IT industry. We contacted the interviewees by email first to prepare for the interviews.

We conducted eight interviews (five telephonic interviews and three face to face interviews). According to the interviewees willing, the interviewees’ names and their companies are anonymous A pilot-test was performed before we did the interviews to estimate the time needed for interviewing each participant and to check if there is anything wrong in the interview guide.

We have mentioned only six interviews in our results because the other two were not accurate (the first developer who we made face to face interview was using another technique than RDFa in structuring his website. The second developer who we made a telephonic interview was planning to use RDFa but did not implement it yet).

Data was collected by writing notes during the interviews except one interview, its data was collected by using mobile recorder (one of the face to face interview); each interview lasted between 25-50 minutes.
3.3 Data Analysis and Presentation

Data analysis consists of different steps. First step is organizing the data for analysis: this involves transcribing the interview data. Second step is reading the data to get overall meanings of data. Third step is beginning detailed analysis. Forth step is generating description. Fifth step is representation of description and last step is making interpretation of the data [17]. Different techniques can be used to analyse interviews ‘data. These techniques can be used to organize the interview text, to condense the interview in the form of some short sentences in order to get the meanings of what was said in the interview. The work of transforming the collected data into an understandable text was very extensive and was carried out in several stages. The first stage was to transcribe all interviews. This text was then processed and shortened in order to complete a first draft of the empirical findings. However this text became very extensive and it was soon realized that if the empirical findings were presented in such a way the reader would have found it very hard to get a reasonable oversight of the data.

We first transcribed the data collected from the interviews and from that transcribed data we constructed a summary of each interview with respect to our interview questions. Then we have presented interviews discussions with respect to the main topics of our empirical study and what are the opinions of our interviewees. From the interviews discussions we have presented our findings in a tabular form. From this table we have extracted the results according to our research questions. We have presented the results in tabular form and in flowchart.

3.4 Research Quality

In order to enhance the credibility of our work, we use methods to enhance the research validity and we put a strong emphasis on the ethical aspect of our research. Both validity and ethics are discussed as follows:
3.4.1 Validity and Reliability

Qualitative validity means to check the accuracy of the findings from empirical study by implementing certain techniques [17]. We apply pattern matching method to ensure validity of our research. Pattern matching is a method which compares empirical collected data with predefined data as in our theoretical part [28]. We are aware of the fact that literature presents a more theoretical point of view, whereas interviews with the web developers will highlight practical aspects. Yet, if both data sources will generate similar outcome, then our research outcome will gain more validity.

Reliability is an important factor in research quality because it examines the consistency and stability of the approach that is used by researchers. There are different procedures to check the reliability i.e. verifying that the transcripts do not have mistakes, there should not be a drift in definition during coding process, in the team research there should be meeting to coordinate the communication among the coder and there should be cross check [17]. We have transcribed the interview’s data very carefully to avoid mistakes.

3.4.2 Ethics

Ethics should be considered especially when any research involves humans and can affect them as well. Kvale [31] highlights three key points to be well thought-out when conducting interviews:

Informed consent: the interviewees have to approve participating in the research, including knowing the subject and purpose of the research, and how their answers will be used [31]. In our case we have informed our subjects beforehand about the purpose of our work and how we are going to deal with their answers.

Confidentiality: during the interview sessions it is often possible to reveal personal details of interviewees. Therefore, they must be informed that their responses will be dealt with full confidentiality [31]. We have chosen to keep our interviewees and their companies’ names anonymous according to interviewees’ request.
Consequences: minimizing the risk of harm to the interviewees by balancing harm and benefits of the research is an overarching principle when conducting a research based on interviews [31]. The type of harm possible during an interview in our work would less likely have a psychological nature (like the intimacy of a therapeutic interview) but rather a possible work-related conflict, if the employer of the interviewee would use his or her answers in a disadvantageous way. It could be the case if the interviewee would accidentally reveal information not intended for the public or if the employer would find out about something that can lead to negative consequences for the interviewee.
Chapter 4. Results

In this chapter of our thesis, we present the results of our empirical study. We have presented these results according to our research questions- What are the benefits and barriers of using RDFa in structuring web sites and enhancing their web visibility? And - How do web developers implement RDFa in structuring web sites?

4.1 Interview Discussions

In this section we discuss the empirical findings that we have extracted from the interview’s summaries, see Appendix 4. First we present interviewees introduction under the coming heading then we present our findings under the main headings of our topic of study.

4.1.1 Interviewees

We have interviewed different web developers from different companies. Our Interviewee 1 is working as a web developer at an audio video solutions related company in Dubai. Interviewee 2 is working as web developer at a video streaming and advertisement Company in Holland. Our interviewee 3 is working as team leader in web development at an electronics related company in UAE. Interviewee 4 is working as software engineer in web development in USA. Interviewees 5 and 6 are working as web developers in different software houses in Pakistan.

4.1.2 Motivation to use RDFa

Our all interviewees have different motivation behind using RDFa in structuring their web sites. Interviewee 1 believed that implementation of RDFa has increased the web traffic to their web sites. According to interviewees 2 and 5, RDFa makes the code machine readable, increase the web visibility and different vocabularies can be used with RDFa. Interviewee 3 thinks that RDFa is a stable, powerful, flexible, W3C recommended, RDF triple can be
used and it based on DRY idea. Interviewee 4 believed that RDFa increases the web visibility and it is W3C recommended. Interviewee 6 was motivated to use the RDFa because he thought that RDFa increases the ranking of the web site in search engine and the usage of RDFa saves money.

4.1.3 Requirements to implement the RDFa

We have asked the question to all interviewees about the implementation requirement of RDFa in structuring of web sites. We got different opinions from each interviewee. Interviewee 1 stated that good planning and understanding of the ultimate objective of implementing RDFa is important and the code should be written in XHTML1.1 or later versions. Interviewee 2 believed that code should be in XHTML1.1 because RDFa can be implemented only in XHTML1.1. According to interviewee 3, code should be shifted to XHTML1.1 to implement the RDFa. Interviewee 4 thought that code should be written in XHTML1.1. Interviewee 5 stated that HTML code should be converted into XHTML in order to implement the RDFa. Interviewee 6 believed that code should be written in XHTML, DOCTYPE should be declared and select the suitable vocabulary to implement the RDFa. All interviewees stated that the main requirement to implement RDFa is that code should be written in XHTML1.1.

4.1.4 Technical steps to implement the RDFa

Our interviewees agreed on this step that code should be written in XHTML1.1. Interviewee 2 added that DOCTYPE contains XHTML+RDFa1.0 but DOCTYPE should be changed to XHTML+RDFa1.1 if vocabularies need to be implemented in web sites and root element must be HTML. Interviewees 3 and 4 stated the same technical steps in the selection of vocabularies but interviewee 3 added the DOCTYPE declaration. Interviewees 5 and 6 explained the same technical steps which are: Root element should be HTML, DOCTYPE should be declared and URIs should be converted into CURIEs.
4.1.5 External Vocabularies

We have found that all interviewees have used the external vocabularies. Interviewees 2 and 6 have used FOAF and GR. Interviewees 4 and 5 have used the FOAF and DC as external vocabularies. Interviewee 1 has used the vocabularies GR and DC. Interviewee 3 has used only GR. FOAF has been used by interviewees 2, 4, 5 and 6. DC was used by interviewees 1, 3 and 5. GoodRelations has been used by interviewees 1, 2, 3 and 6. We noticed that most of the interviewees have used FOAF and GoodRelations.

4.1.6 Barriers of RDFa

There are some barriers to implement the RDFa in structuring of web sites. We mention here according to interviewees what are the barriers of RDFa. Interviewees 1, 2 and 3 stated web sites should be written in XHTML 1.1. Interviewee 4 said the same thing but in another way; he said that code can not be written in HTML because RDFa can not be implemented in HTML. Interviewees 5 and 6 stated the same barriers like: code should be written in XHTML1.1, XHTML cleaning tools can not be used because it affects the RDF triples and URIs should be converted into Curies. According to all interviewees the main barrier in using RDFa is that code should be written in XHTML1.1.

4.1.7 Semantic markup and web visibility

We got different views of interviewees about web visibility. According to Interviewee 1 they have used semantic markup in their web site and web visibility of their web site has increased in search engines. Interviewees 2 stated that RDFa makes the code machine readable so the search engine’s software understands the code. As a result website gets high ranking in search engine. Interviewee 3 explained that his company has used RDFa in their web site and they got significant results i.e. web site is more visible, web traffic increased to their web site and they are first in results in search engines. According to interviewee 4, semantic markup adds semantic to web site and makes the code machine readable. The browser can understand the contents of the web site and this leads to increase the web
visibility in search engines. Interviewees 4 and 5 stated that semantic markup makes the code machine readable and it increases the web visibility in search engines. All interviewees mentioned that semantic markup increases the web visibility because semantic markup makes the code machine readable.

4.1.8 Evaluation of web visibility

While interviewing the web developers we have noticed that all developers used the Google Analytics to measure the web visibility of their web sites. Only interviewees 4 and 5 used other tools also. Interviewee 4 used search test on different search engines to measure the web visibility. Interviewee 5 used different tools but more commonly he used Google Analytics.

4.1.9 Benefits of RDFa

All interviewees stated the different benefits of RDFa. According to interviewee 1, RDFa is easy to implement and maintain and it increase the ranking of web site in search engines. Interviewee 2 stated, RDFa makes the code machine readable, RDF triples can be used with RDFa and different vocabularies can be used. According to interviewee 3 RDFa is easy to use, RDF triple can be used, external vocabularies can be used and RDFa makes code machine readable. Interviewee 4 mentioned that RDF triple can be used with RDFa, vocabularies can be used and RDFa used XHTML attributes as well as its own attributes. Interviewees 5 and 6 stated the same benefits of RDFa like it is easy to use, it makes the code machine readable, different vocabularies can be used and it increases the web visibility. In addition interviewee 6 mentioned that RDF triple can be used with RDFa and implementation of RDFa saves money.

4.1.10 Difficulties of RDFa

Our interviewees have faced difficulties during implementation of RDFa in structuring of web sites. Interviewee 1 thought that it was difficult to redesign the web site. Interviewee 2
stated that it is difficult to change the code in XHTML1.1 and convert URIs into CURIEs. Interviewee 3 faced the same difficulty as interviewee 1 i.e. redesigning the code. Interviewee 4 thought that it is a time consuming process to redesign and optimize the code. Interviewee 5 did not face any difficulty in implementing RDFa. Interviewee 6 stated that it is difficult to convert the code in XHTML1.1.
We have summarized our finding from the interviews discussions in Table 1.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Interviewee 1</th>
<th>Interviewee 2</th>
<th>Interviewee 3</th>
<th>Interviewee 4</th>
<th>Interviewee 5</th>
<th>Interviewee 6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Motivation to use RDFa</strong></td>
<td>Increase their web visibility by increasing the traffic to their web site.</td>
<td>The code readable, well structured and descriptive. Different vocabularies can be implemented.</td>
<td>RDFa is stable, flexible and powerful. W3C recommended using RDF triples and it’s easy to implement</td>
<td>RDFa is reliable, increase the web site visibility and it’s a W3C recommended.</td>
<td>RDFa make the code machine readable. Increased the web visibility. External Vocabularies used</td>
<td>RDFa increased the ranking of web site in search engine. Saves money.</td>
</tr>
<tr>
<td><strong>Requirements to implement RDFa</strong></td>
<td>Good plan and understanding required. Code must be written in XHTML1.1.</td>
<td>The code must be written in XHTML 1.1 or further version.</td>
<td>Shift the code to XHTML 1.1.</td>
<td>The code must write in XHTML1.1.</td>
<td>HTML code should be converted into XHTML1.1.</td>
<td>Code should be written in XHTML1.1. DOCTYPE declaration. Vocabularies selection</td>
</tr>
<tr>
<td><strong>Technical steps to implement RDFa</strong></td>
<td>Redesign and restructure the web site in XHTML 1.1 and select the right vocabulary.</td>
<td>The code written in XHTML1.1 DOCTYPE contain XHTML+RDFa 1.0. DOCTYPE must change to XHTML+RDFa 1.1 if external vocabularies required implementing.</td>
<td>The code should be in XHTML1.1. DOCTYPE declaration and selection of vocabularies</td>
<td>The code must be in XHTML 1.1 to make the adding of own vocabulary or other vocabulary easier and more efficient.</td>
<td>Root element HTML. DOCTYPE declaration. Code in XHTML1.1. CURIEs</td>
<td>Root element HTML</td>
</tr>
<tr>
<td><strong>Vocabularies used</strong></td>
<td>GR and DC</td>
<td>FOAF and GR</td>
<td>GR</td>
<td>FOAF and DC</td>
<td>FOAF and DC</td>
<td>FOAF and GR</td>
</tr>
<tr>
<td><strong>Barriers in using RDFa</strong></td>
<td>Redesigning the code in XHTML 1.1</td>
<td>It can not be used in HTML</td>
<td>The code must be in XHTML1.1</td>
<td>The code can not be written in HTML</td>
<td>Code in XHTML 1.1 XHTML Cleaning tool can not be used. URIs to CURIEs</td>
<td>Code in XHTML 1.1 XHTML Cleaning tool can not be used. URIs to CURIEs</td>
</tr>
<tr>
<td><strong>How semantic markup enhance web site visibility?</strong></td>
<td>Semantic markup helps search engines to find their web sites easier.</td>
<td>The code is machine readable so the web is more visible.</td>
<td>The code is more machine readable so the web site is more visible.</td>
<td>RDFa adds semantic to the web site’s code.</td>
<td>RDFa make the code machine readable.</td>
<td>RDFa make the code machine readable.</td>
</tr>
<tr>
<td><strong>Evaluate web site visibility</strong></td>
<td>Google analytic</td>
<td>Google analytic</td>
<td>Google analytic</td>
<td>google analytics by doing test searches search engines</td>
<td>Google analytic</td>
<td>Google analytic</td>
</tr>
<tr>
<td><strong>Benefits of semantic markup</strong></td>
<td>Increased ranking of web site in search engines Easy to implement and maintain</td>
<td>The code is machine readable, understandable by the browsers. RDF triples and different vocabularies can be implemented,</td>
<td>Increased the web visibility in search engines.</td>
<td>Increased the ranking in search engines.</td>
<td>Easy to use. Make the code machine readable. Vocabularies can use. Increased the web visibility.</td>
<td>Easy to use. Make the code re usable Different Vocabularies can use. RDF triples can use. Save money</td>
</tr>
<tr>
<td><strong>Difficulties of RDFa</strong></td>
<td>Difficult to redesign.</td>
<td>Change the code to XHTML 1.1. CURIEs.</td>
<td>Redesign the code in XHTML1.1</td>
<td>It takes him/her long time to redesign the web site</td>
<td>Not mention</td>
<td>Hard to convert the code into XHTML</td>
</tr>
</tbody>
</table>

**Table 1:** The interview’s findings
4.2 Interviews Results

4.2.1 Benefits and Barriers of Using RDFa

In this section we have summarized the results that we have found from our empirical study. As we have mentioned in our goal and expected outcome that this thesis will provide benefits and barriers of using RDFa, and the guideline to use RDFa. Table 2 shows the benefits and barriers of RDFa.

<table>
<thead>
<tr>
<th>Benefits of using RDFa</th>
<th>Barrier and Difficulties of using RDFa</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDFa easy to implement because it is based on DRY (Don’t Repeat Yourself) idea.</td>
<td>RDFa do not support HTML code must be written in XHTML</td>
</tr>
<tr>
<td>Supported by W3C</td>
<td>Convert the URIs into CURIEs</td>
</tr>
<tr>
<td>Different Vocabularies can be implemented with RDFa like FOAF, GoodRelations and DC</td>
<td>XHTML cleaning tools can not be implemented</td>
</tr>
<tr>
<td>RDFa make the code machine understandable</td>
<td></td>
</tr>
<tr>
<td>RDF triples can be implemented in RDFa</td>
<td></td>
</tr>
<tr>
<td>It is easy to maintain the web site structured in RDFa</td>
<td></td>
</tr>
<tr>
<td>RDFa enhanced the web visibility of web sites in search engines because it makes the code machine readable</td>
<td></td>
</tr>
<tr>
<td>Implementation of RDFa saves money</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Benefits and barriers of using RDFa
We have summarized the benefits and barriers of RDFa with respect to each interviewee in Table 3 and Table 4 respectively.

<table>
<thead>
<tr>
<th>Interviewees</th>
<th>Vocabularies</th>
<th>Increase web visibility</th>
<th>The code machine readable</th>
<th>Easy to implement</th>
<th>W3C recommendation</th>
<th>Low cost</th>
<th>RDF triples used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Benefits of RDFa with respect to each interviewee

<table>
<thead>
<tr>
<th>Interviewees</th>
<th>Code written in XHTML 1.1</th>
<th>Covert URIs into CURIEs</th>
<th>XHTML cleaning tool can not be used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Barriers of RDFa with respect to each interviewee

4.2.2 Guideline

In this section we suggest a Guideline for companies that want to use the RDFa in structuring their web sites in the light of our empirical study. This guideline is based on the answer to our second research question i.e. how do web developers implement RDFa in structuring web sites? Figure 16 illustrates this guideline in the form of a flowchart.

The process of generating the guideline

We have extracted the steps of our guideline from the empirical study and the literature review which have been listed below:

1-We have selected the keywords from our empirical study like: XHTML1.1, Vocabularies, DOCTYPE, Root element, CURIEs and web visibility. These keywords have been selected from the start of the study and the keywords based on our own knowledge and experience.
2- The next step is based on what our interviewees had said about these key words that we have selected in the previous step. During the questions about technical steps and requirement to implement RDFa, each interviewee had said that the code should be written in XHTML1.1 or if the code has been written in HTML then it should be converted into XHTML1.1. From our literature review we have found that RDFa can only be implemented in XHTML 1.1 or later versions. This step has been stated as the first step of our guideline because this is the main requirement to implement RDFa.

3- Next step of the guideline is the selection of vocabularies. We have generated this step according to the key word “vocabularies”. Each interviewee has used the vocabularies and as we have mentioned in our literature review about the different vocabularies like GoodRelations, FOAF and Dublin core according to their requirement or need. This explains that these vocabularies have different properties fitting different purposes and every company has their own requirement which decides the choice of vocabulary.

4- Next steps based on the key word DOCTYPE the declaration of which is based on the choice of vocabularies. To use the vocabularies it is vital to declare the right DOCTYPE.

5- Root element should be HTML; we took it as the next step in our guideline because all XHTML documents must have HTML as root element.

6- The next step is based on the key word CURIEs. We have selected this step after all the fundamental steps because at this level the implementation of RDFa has been started.

7- The last step of the guideline is the implementation and evaluation. This step is based on the key word “web visibility”. We have selected this step as the last step because after implementation of RDFa there is a need to evaluate the effect of RDFa in the web visibility of the web site in search engines.
Guideline steps

Guideline steps has ordered according to the logical sequence of implementing RDFa e.g. RDFa can not be implemented if the code was not written in XHTML1.1; vocabularies can not be used if they were not selected. We have explained the guideline under the following steps:

The code should be written in XHTML 1.1

According to our empirical study and literature review RDFa does not support HTML because it is not extensible. XHTML is an extensible version of HTML. XHTML 1.1 or later versions are required to implement the RDFa. For more details about XHTML see Appendix 3. This is the first step in the guideline because it is the main requirement to implement RDFa. When a company wants to implement the RDFa in their web site then it is required to write the code in XHTML1.1 or converts the code if it has been written in HTML.

Selection of Vocabularies

RDFa can use the different external vocabularies and that is the main benefit of RDFa. Selection of the vocabularies based on the companies’ requirements. In the suggested guideline to implement the RDFa, after converting the code into XHTML it is important for the developers to select the vocabulary according to their needs. Developers can choose different vocabularies according to the requirements of the web sites. For example if the web site is related to e-commerce then the GR vocabulary can be used. GoodRelations is ontology of e-commerce. GoodRelations can be used for different purposes like business information, product information, prices information etc. If the web site is about persons and their relationship then FOAF could be a suitable choice for web developers to describe their resources. FOAF is a vocabulary about the persons and their relations. In case, the web site requires information about the things like references, locations, companies then DC vocabulary could be a suitable choice for the developers. DC is a vocabulary about describing the resources. There are many other vocabularies can be used for different
purposes but the above mentioned vocabularies has been used by our interviewees and these are the prominent vocabularies in the current market.

**DOCTYPE Declaration**

DOCTYPE should be declared in the header as XHTML+RDFa1.0 to implement the RDFa in structuring the website. By declaring the DOCTYPE, the attributes of RDFa and XHTML can be used together.

**Usage of Vocabularies**

With RDFa, external vocabularies can be implemented into the website’s structure but the DOCTYPE should be changed as XHTML+RDFa1.1. In RDFa multiple vocabularies can be used in a single page. RDFa could be either hard coded into the website or generated dynamically form the databases with help of the external vocabularies. RDFa attributes (like typeof, about and property etc.) give the opportunity to implement different external vocabularies in the structuring of websites [5]. For example, if it is required to use the GR or FOAF vocabulary in structuring of websites. GR can be used in the following steps:

1- Adding company information by using GoodRelations with RDFa. First thing to be done after declaring the DOCTYPE is converted the URIs to CURIEs in the Div tag. See Figure 13.

```
<div xmlns:GR="http://purl.org/goodrelations/v1#">
```

**Figure 13**: Converted URI to CURIE

2- With help of RDFa, company’s information can be added. See Figure 14.

```
<div about="#company" typeof="GR:BusinessEntity">
<div property="GR:legalName" content="Honda Atlas Cars Ltd."/>
```

**Figure 14**: Usage of GoodRelations for adding company information
Same steps mentioned above can be repeated if it is required to add personal information by using FOAF. See Figure 15.

```xml
<div xmlns:FOAF="http://xmlns.com/foaf/0.1/">
  <div typeof="FOAF:Person">
    <span property="FOAF:name">Maria</span>
  </div>
</div>
```

**Figure 15:** Usage of FOAF for adding personal information

### Root Element

The Root element of the local part of document must be HTML:

```html
<html xmlns = http://www.w3.org/1999/xhtml xml:lang ="en">```

### Convert URIs to CURIEs

Normal URIs should be converted into CURIEs. In normal URIs it is required to write the complete URIs but in CURIEs the main part of the URIs is assigned to a token. To create the CURIE, the first step is prefix mapping that link the main part of the URI to a prefix. In RDFa prefix mapping is written in XML namespace syntax. When the prefix has been established then the developer can short the URIs. In the processing of RDFa the CURIEs is resolved to their original URIs.

### Implementation and Web sites’ Evaluation

We recommend the web developers to evaluate their web sites after implementing RDFa by using an analytic tool or by making searching test to their web site in different search engines. There are different analytic tools are available in the market but most developers use the google analytic as shown in our empirical study. The benefits of using an analytic tool are that it helps developers to identify the weak points in the page structure and the amount of web traffic to their web site. In the other words, it can evaluate the web visibility of web site in search engines by using the analytic tools. On the other hand, making searching test to the web site in different search engines help the developers measure the
ranking of their web site in search engines. A developer can see the effect of using RDFa in his web site related to web visibility.
Figure 16: Flowchart of the guideline to implement RDFa
4.3 Summary of Results

In this section we discuss the results of our empirical study on the basis of our interviews that we have conducted with different web developers from different companies. Our discussion is based on the views of our interviewees on key points: motivation behind using of RDFa, requirements to implement RDFa, technical steps to implement RDFa, barrier and benefits, web visibility and vocabularies used with RDFa. Our interviewees have used RDFa in structuring their web sites. They have different opinions about the implementation of RDFa in web sites.

We have divided our results into two parts. The first part of the results is the benefits and barriers of using RDFa and this part has been represented in Table 2. The second part of our results is a guideline for the companies that want to implement the RDFa as semantic markup technique in structuring their web sites. According to our interviewees RDFa is easy to use. Our interviewees mentioned that RDFa is supported by W3C and that is benefited. Different vocabularies can be implemented with RDFa in structuring of web sites, all interviews have mentioned about this benefit. All interviewees were agreeing that RDFa make the code machine understandable. Some of our interviewees believed that RDF triples can be used with RDFa and there are no special rules to implement the RDFa. All the interviewees agreed that semantic markup enhance the web visibility in search engines because it makes the code machine readable.

We have extracted the barriers of RDFa as we have presented in Table 2. Our all interviewees were agreeing on that, the main barrier in the implementation of RDFa is to convert the code into XHTML if the code has been written in HTML. In addition, some of our interviewees have mentioned more barriers like XHTML cleaning tool can not be used because it affect the RDF triples and URIs should be converted into CURIEs.

The second part of our results is a guideline. We have created this guideline in the light of our empirical findings. We have asked to each interviewee about requirements and technical steps for implementing RDFa in structuring web sites. Firstly, we have described the guideline steps and then we have created a flowchart of the guideline see Figure 16. According to our interviewees, the main barrier is that RDFa can not be used in HTML.
The companies that want to implement RDFa in their web sites should write the code in XHTML. In case, if the companies want to use external vocabularies with RDFa, then they have to change the DOCTYPE in the header of their web page into XHTML+RDFa1.1. Furthermore, root element should be HTML. Finally, the URIs should be converted into CURIEs.

4.4 Conclusion

We have concluded after the results and the results discussion that RDFa increases the web visibility of web sites in search engines. RDFa makes the code more machine understandable, easy to implement, supported by W3C, different vocabularies can be implemented with RDFa, RDF triples can be used in RDFa and there are no special rules to implement the RDFa. There are few barriers in using RDFa, like the code should be written in XHTML, XHTML cleaning tools cannot be used and it is required to convert the URIs into CURIEs. In order to implement RDFa, developers need to write the code in XHTML, declare the DOCTYPE in the header according to their requirements and the root element should be HTML.

4.5 Threats to Validity

In this section we present the possible threats to our study. Firstly, the participants know the interview guide questions before we do the interview. The threat is that the participants prepare their answers in advance. In order to avoid this threat, we didn’t send our interview guideline in advance but we sent summary of what our thesis is about, and we asked our questions through the interviews. Secondly, the wording of questions through the interviews could be a possible threat to construct validity. To minimize this threat, we were eager to formulate clear and uncomplicated questions to get the desired answers from the participants. To minimize the threat to external validity, we chose different developers from different companies. Finally, the major threats to our study are the size of sample of recruited participants to the interviews and the two different types of interviews we made (telephonic and face to face interview). The difficulty of recruiting more participants since RDFa is a new technology made us accept these two threats.
Chapter 5. Discussion

In this chapter we have discussed our study in the light of our theoretical baseline and the empirical study. First we repeat our research questions i.e. what are the benefits and barriers of using RDFa in structuring web sites and enhancing their web visibility? And how do web developers implement RDFa in structuring web sites? To answer these questions we have collected our secondary data from books, journals and articles. We have conducted the empirical study to collect the primary data to answer our research questions. Now we will discuss our finding with respect to theoretical baseline and empirical study.

Benefits of using RDFa

We have found the benefits of RDFa in our literature review in section 2.3.6 from different sources. According to Graf [1] RDFa makes the code human and machine readable, and we have found the same result in the empirical study. Interviewees 2, 3, 5 and 6 have the same opinion i.e. RDFa makes the code machine readable. According to our interviewees RDF triples can be used with RDFa. According to Adida [4] RDF triples are not coupled with RDFa, it can be copied and pasted. This make the RDF decoupled from the code.

According to Graf [1] different vocabularies can be implemented with RDFa, and it is easy to use because RDFa uses the XHTML attributes as well as its own attributes also. It makes it easier to extract the RDF triples. Same results we have found from our interviewees i.e. external vocabularies can be implemented with RDFa and it is easy to implement the RDFa. All interviewees have used the vocabularies like GR, FOAF and DC. RDFa supported by W3C. W3C is the main international organization that works on the World Wide Web. Our interviewees 3 and 4 have mentioned that RDFa is supported by W3C.

According to Thomas [9] RDFa is based on the DRY idea, every piece of the code is written with all information only one time, in order to avoid repeating the same code. It makes the code reusable. Our interviewees have mentioned the same benefit. According to Graf [1] there are no special rules to implement the RDFa, a web developer can use their
own standards. In our empirical study, we have found the same result from our interviewees that there are no special rules to implement the RDFa.

**Barriers of using RDFa**

As we have mentioned the barriers of RDFa under the heading 4.2.1, the main barrier of the RDFa is that the page content must be written in XHTML 1.1 or later versions of XHTML. According to Adida and Birbeck [3] it is not possible to implement RDFa in HTML because it is not extensible language. During our empirical study we have got the same result from our interviewees. All of our interviewees were agreeing on that the code should be converted into XHTML in case it is written in HTML in order to implement the RDFa in structuring the web sites.

Graf [1] explained another barrier that cleaning tools can not be used in XHTML to make the page well formed because it affects the RDF triples. Our interviewees 5 and 6 have mentioned that they faced the same barrier through implementing RDFa in structuring their web sites. Our all interviewees have faced that converting URIs to CURIEs is another barrier in using RDFa in the web sites structure. As it is mentioned in section 2.3.6 CURIEs are special type of URIs. According to W3C, the URIs must be changed to CURIEs. In CURIEs the leading part of URIs assigned with a token.

These are the benefits and barriers of RDFa. In this discussion we have compared our findings from the empirical study to studies that have been done by others. In the next section of our discussion we will discuss the guideline that we have extracted from the technical steps of implementing RDFa. Through our theoretical study we have not found this kind of guideline done before.

**Guideline to implement RDFa**

We have created a guideline for the companies that want to use RDFa in structuring of their web sites. This guideline is based on our empirical study. During the empirical study we have asked questions to all of our interviewees that “what are the requirements to implement RDFa for structuring the website?” and “what are the technical steps to implement the RDFa semantic markup in structuring of websites?” We got different answers from each interviewee. We expect this guideline will be useful to developers who
want to implement the RDFa. This guideline may help developers to minimize the barriers of RDFa.

The first step of the guideline is to write the code in XHTML 1.1 because RDFa does not support HTML. It has been mentioned in section 2.3.6 that the page contents should be written in XHTML 1.1 or later version. The second step of the guideline is DOCTYPE declaration. DOCTYPE should be declared in the header as XHTML+RDFa1.0 see the DOCTYPE declaration in Figure 8. DOCTYPE should be XHTML+RDFa1.1 if the external vocabularies like GR, FOAF and DC need to implemented with RDFa in structuring web sites. Third step of the guideline is the root element should be HTML. In Figure 8 it is shown that root element should be HTML. The forth step is to convert the URIs into CURIES.

**Web visibility in Search Engines**

In our empirical study we have asked a question to our interviewees that How semantic markup enhance the web visibility of websites in search engines? RDFa is a semantic markup technique and our focal point during the empirical study was the RDFa. According to our interviewees RDFa makes the code machine readable so it is understandable by the crawler during the search engine process, as a result the visibility of web sites in the search engine increased. SEO and search engine process has been discussed in section 2.2.1. It has already been mentioned in section 2.3 that semantic markup makes the code machine readable. Computer can understand the code, as a result web sites get higher ranking in the search engines.

In this chapter we have discussed our results from empirical study in the light of our literature reviews. This discussion shows the validity of our results because we have compared the empirical result with the literature review that we have found from different articles, books and other online sources.
Chapter 6. Conclusion and Future Work

The topic of this thesis is semantic markup, and more specially RDFa. In semantic markup we have focused on the RDFa. We have raised two research questions related to RDFa i.e. what are the benefits and barriers of using RDFa in structuring web sites and enhancing their web visibility? And how do web developers implement RDFa in structuring web sites? To answer these questions we conducted interviews with web developers using the RDFa in structuring their web sites.

We have examined the benefits and barriers of RDFa in practical field by conducting the interviews. We have also constructed a guideline, created for the companies that want to implement the RDFa as a semantic markup technique in structuring of web sites. We did not find this kind of guideline during the literature review, so we have found that this is an advanced research in RDFa as a semantic markup technique.

We have found almost the same benefits and barriers in the interviews as found in the literature review described in section 2.3.6. In a sense we have conformed the benefits and barriers of using RDFa during our empirical study. The guideline to implement RDFa as semantic markup technique in structuring web sites is a new work. We have presented the guideline as a flowchart of the involved steps that would help web developers to implement RDFa.

Future Work

In this thesis we have created a guideline for the companies that are planning to implement RDFa in their web sites. This guideline that we have created needs to be evaluated by sending it to the interviewees to ask their opinion. Another kind of evaluation is to ask web developers to use the guideline in different contexts; due to time constraints we could not evaluate the guideline. The guideline could be improved after the evaluation. The future work of our thesis could be an investigation of how search engines use RDFa in ranking process. One more suggestion is to carry out a survey to get more insight into the topics covered in the interviews.
References


http://books.google.se/books?id=yraHOhPTxEOC&printsec=frontcover&dq=xhtml&hl=sv&ei=6fRcTcmGHYav4AbG3Y2pCw&sa=X&oi=book_result&ct=result&resnum=5&ved=0CEMQ6AEwBA#v=onepage&q=&f=false [Feb. 17, 2011]


[29].”RDFa in XHTML: Syntax and Processing”, [Online]: http://www.w3.org/TR/rdfa-syntax/ [May. 27, 2011]


[40].”XHTML 1.0”, [Online]: http://www.w3.org/TR/xhtml1/ [Feb. 18, 2011]

[41].”XHTML Syntax”, [Online]: http://www.w3schools.com/xhtml/xhtml_syntax.asp [Feb. 18, 2011]
Appendix 1: Interview Guide

Date and time:  ---------------------------------------------------------

Interviewer:  ---------------------------------------------------------

Interviewee  ---------------------------------------------------------

Company:  ---------------------------------------------------------

----------------------------------------------------------

----------------------------------------------------------

Purpose of the Interview:

The purpose of the research is to investigate the current work in semantic markup. Within the Semantic Markup our focus is on RDFa. As our research question is how does semantic markup enhance website visibility in search engine? The goals of our research are to investigate the benefits of semantic markup and to create a guideline for helping companies to structuring their web sites semantically to increase the web visibility in search engines.

Abbreviations:

RDFa=> Resource Description Framework attribute

FOAF =>Friend of a Friend

GR=> GoodRelations

DC=>Dublin Core
Questions:

1- Do you use RDFa as semantic markup in structuring of your website? If you don’t, why?

2- What are the factors that motivate you for selecting RDFa in structuring of your website? In other words, why are you using RDFa in your website?

3- According to your experience what are the requirements to implement RDFa for structuring the website?

4- In your opinion, what are the technical steps to implement the RDFa semantic markup in structuring of websites?

5- Which vocabulary (FOAF, DC, GR etc.) have you used with RDFa in structuring your website?

6- In your experience, are there any barriers in using RDFa in structuring the website? If so, what are these barriers?

7- How does semantic markup enhance the web visibility of websites in search engines?

8- How you evaluate your website visibility in search engines?

9- According to your experience, what are the benefits of semantic markup?

10- Did you face any difficulties in implementing RDFa in your website?
Appendix 2: Abbreviations and their Definitions

SEO=> Search Engine Optimization

RDF=> Resource Description Framework

RDFS=> Resource Description Framework Schema

RDFa=> Resource Description Framework in attribute

XHTML=> Extensible Hyper Text Language

W3C=> World Wide Web Consortium

XML=> Extensible Markup Language

HTML=> Hyper Text Markup Language
Appendix 3: XHTML

Markup gives more information about the document by adding information that explains how the document is structured and presented. Markup is available in most of the web pages that we deal with everyday. We may not notice that because all we see is on our computers’ screens is a page of text, while all the formatting and structuring of these web pages was done by Markup. XHTML is one of the markup languages that used to tell the browsers about how the web page structured and what is the way to display it [2].

XHTML stands for Extensible Hyper Text Markup Language. It is cleaner version of HTML, and it is created by the combine features of HTML and XML (Extensible Markup Language). It is a recommendation of W3C [36]. The web sites are more machine readable if they restructured in XHTML. Browser can understand much better the contents of the web site developed in XHTML as compared with web sites that are developed in HTML [36].

XHTML was invented to extend subset and reproduce HTML 4. XHTML based totally on XML. It is specially designed to work with user agents that based on XML, and all XHTML’s documents being:

- Viewed, read, validate and edit by same standard tools that being used in XML.

- Can be written by XHTML 1.0 conforming user agents to work at the same level or better than if they are written by HTML 4 conforming user agents.

- Achieve applications e.g. scripts and applets by using the XML Document Object Model (DOM) or the HTML Document Object Model.

- Likely working inside and among different XHTML environments.

By using the XHTML the developers can take all benefits of XML plus the XHTML abilities. For all this above, XHTML is the future of the internet development [40].
XHTML Syntax rules:

- All attribute in XHTML must be written in small letter.

  <table width=“100%”>

- The attribute values must be written between two quotations.

  <table width=”100%”>

- It’s not allowed to minimize the attribute in XHTML.

  <input checked=”checked”/> not allowed to be minimized to <input checked>

- There is a lang attribute in XHTML which can be used in most of XHTML elements to specify the language of the element’s content.

  <div lang=”it” xml : lang=”it”>Ciao bella!</div>

- There are some compulsory elements in XHTML, these elements have to have a DOCTYPE declaration. The elements that have to be presented are html, head, title and body.

  <!DOCTYPE doctype goes here>

  <html xmlns=”http://www.w3.org/1999/xhtml”>

  <head>

  <title> Title goes here</title>

  </head>

  </body>

</html>

- An XHTML document has to contain three main parts DOCTYPE declaration, <head> and <body>. [41]
Appendix 4: Interviews’ Summaries

In this section the summarizations of the interviews with respect to interviewees and questions that we have asked during interviews have presented.

Interviewee 1:

The first interviewee works as a web developer in an Audio and Video Solutions Company in Dubai, UAE.

- **Motivation behind using RDFa**

He said that his company decided to implement RDFa in structuring their company’s web site. They wanted a technology to give their web site a new dimension of visibility and to have more control on how they want their web site to be found by search engines, which led to increase web traffic to their web site and to target a bigger group of people.

- **Requirements to implement RDFa**

Interviewee 1 thought that there are many requirements to implement RDFa for structuring web sites. The most important requirements are a good plan and a good understanding of what exactly needs to be reached by implementing RDFa. Furthermore, the code must be well designed and written in XHTML 1.1 or later versions.

- **Technical steps of implementing RDFa**

In his opinion the technical steps to implement RDFa as semantic markup in structuring of web sites. Firstly, translate the needs of using semantic markup into technical specifications. Secondly, redesign and restructure the web site in XHTML 1.1 as he used or further version. Finally, select the right vocabularies.

- **Vocabularies**

He used in his web site DC and GoodRelations vocabularies.
- Barriers of RDFa

He thought that the only barrier they faced in using RDFa is that they had to redesign the web site in XHTML1.1 because web site was written in HTML.

- How semantic markup enhanced the web visibility

Interviewee 1 said that semantic markup helps the search engines to find their web site easily. According to him/her, since they used semantic markup last year in structuring their web site, the visibility of their web site in search engines increased significantly.

- Evaluation tools that used

He mentioned that he is using google analytics to evaluate the web site visibility in search engines.

-Benefits of RDFa

In his opinion, there are many benefits of semantic markup. The web site gets better rank in search engines; it’s easy to implement and to maintain.

- Difficulties in implementing RDFa

He thought that there are no difficulties in implementing semantic markup except the difficulty of implementing the RDFa in the existing web site because they had to redesign the web site.

Interviewee 2:

The second interviewee works as a programmer and web developer in a company for video streaming and advertisement services in Rotterdam, Holland.

- Motivation behind using RDFa

He said that he was being motivated to use RDFa in structuring his web site because of many factors. The first factor is with RDFa the code is machine readable as well as human readable. The second factor is the code with RDFa is well structured and well descriptive.
The most important factor is, it gives semantic touch to the web site and it is easy to be implemented plus with RDFa an external vocabularies can be used.

- **Requirements to implement RDFa**

According to him/her, before implementing RDFa, the code of the web page should be written in XHTML because RDFa can implement in XHTML 1.1.

- **Technical steps of implementing RDFa**

He said that there are more important technical steps must be done to implement RDFa, in addition to the code must be written in XHTML 1.1, there should be DOCTYPE which contain XHTML+RDFa 1.0., if some vocabularies implemented so the DOCTYPE must be change to XHTML+RDFa 1.1., the root element should be html and should contain XHTML namespace declaration.

- **Vocabularies**

He used in his web site FOAF and GoodRelations vocabularies.

- **Barriers of RDFa**

In his opinion, the limitations of using RDFa are, it can not be used in HTML and the code must be written in XHTML 1.1 or further versions.

- **How semantic markup enhance web visibility**

According to him/her, when semantic markup being used and especially RDFa, the code is machine readable so the search engine software understands the code. As a result for that the web site gets higher ranking in search engine and the web site provide structured information.

- **Evaluation tools that used**

He mentioned that he is using google analytics to evaluate the web site visibility in search engines.

- **Benefits of RDFa**
Pursuant to him/her, there are many benefits of using semantic markup like, it makes the code machine readable which helps to be more understandable by the browser and the RDF triples and different vocabularies can be used in RDFa.

- **Difficulties in using RDFa**

Interviewee 2 thought that there are some difficulties also in implementing RDFa because he has to change the web site’s code to XHTML 1.1 because it was written in HTML before. One more difficult thing is that RDFa needs special type of URIs which called CURIEs.

**Interviewee 3:**

The third interviewee works as a web developers’ team leader in a big international electronics company in Abu Dhabi, UAE.

- **Motivation behind using RDFa**

Interviewee 3 said that they were using microformats as semantic markup but they decided to start using RDFa instead of microformats because it’s more stable, flexible and powerful than microformats. Further more, RDFa based on reliable ontologies which some of them are being for years in the market, and RDFa is a W3C recommendation. All these reasons motivated our interviewee to use RDFa in structuring his web site.

- **Requirements to implement RDFa**

He said that it was a challenge to shift all their code of the web site to XHTML 1.1 as RDFa required to be implemented.

- **Technical steps of implementing RDFa**

Technically, he thought that there are some necessary steps must be considered, like the code must be in XHTML, taking care of the DOCTYPE and namespaces in the code which suites to be used with RDFa, the CURIEs (the compact URIs) and the suitable vocabulary which fits your goal of using RDFa.

- **Vocabularies**
He mentioned that they used GoodRelations vocabulary with RDFa in structuring their web site because he thought it’s matching his goals. As long as, the GoodRelations is ontology for e-commerce and he has an e-commerce web site which wanted to implement RDFa in it.

- **Barriers of RDFa**

According to interviewee 3 the main barrier of implementing RDFa is the code should be written in XHTML1.1.

- **How semantic markup enhance web visibility and the used**

According to him/her, the results of using RDFa in structuring their company’s web site were significant. In few months their web site is more visible on the net. According to a usual evaluation he makes by using google analytic evaluation tool. Google analytic shows that the search traffic increased to the web site from search engines, plus they are in the first results in search engines. He said all this reflected positively on the company’s e-commerce transactions.

- **Evaluation tools that used**

Interviewee 3 used the google analytics to measure the web visibility.

- **Benefits of RDFa**

In his opinion RDFa is a successful technology because it has many benefits like it gives the opportunity to use the RDF triples, the external ontology in structuring the web site which both leads to make the code machine readable. In general, RDFa is easy to implement because it based on DRY idea (don’t repeat yourself).

- **Difficulties in implementing RDFa**

Interviewee 3 faced the difficulties during redesigning the code in XHTML.
**Interviewee 4:**

The fourth interviewee works as Software engineer in a web developing company in Texas, USA.

- **Motivation behind using RDFa**

Interviewee 4 said that he used RDFa in structuring his web site because it’s reliable enough to increase the web site visibility, by giving the opportunity to use the RDF tipples. As well as, RDFa is a W3C recommendation, so the web site’s code has to be a W3C valid and to be written in XHTML before RDFa being implemented.

- **Requirements to implement RDFa**

According to the interviewee 4 code of the web site should be written in XHTML 1.1 because it is required to implement the RDFa.

- **Technical steps of implementing RDFa**

Technically, he thought that it’s very important to carefully design and structure the code of the web page in XHTML 1.1, which make the operation of adding own vocabulary and other external vocabularies easier and more efficient. According to Interviewee 4 Code should be written in XHTML1.1 because it is extensible as compared to HTML. DOCTYPE should be changed to XHTML + RDFa1.1 if external vocabularies need to implement. CURIEs should use instead of URIs

- **Vocabularies**

Interviewee 4 used FOAF and DC vocabularies in structuring his web site.

- **Barriers of RDFa**

Interviewee 4 thought that the code can not be written in HTML because RDFa can not be implemented in HTML.

- **How semantic markup enhances web visibility**
Interviewee 4 stated that RDFa adds semantic to the web site’s code and it makes the code machine readable. The browser can understand the contents of the web pages. All of this leads to increase the web visibility in search engines.

- **Evaluation tools that used**

Interview 4 evaluates his website visibility in search engines by measuring the pagerank of his page by using google analytics tool and search tests using by several search engines.

- **Benefits of RDFa**

According to him/her, RDFa adds semantic to the web site’s code by using RDF triples, own vocabulary and external vocabularies. Furthermore; it makes use of the attributes that already available in XHTML plus its own attributes.

- **Difficulties in implementing RDFa**

According to him/her, there are some difficulties in implementing RDFa like sometimes it takes long time to redesign the web site and optimize it to be ready for implementing RDFa.

**Interviewee 5:**

The fifth interviewee works as web developer in a USA-based software house in Pakistan.

- **Motivation behind using RDFa**

Interviewee 5 used RDFa in structuring the websites. He told the reasons behind using of RDFa are: code become more machine-readable as well as human-readable by using the RDFa, it increased the web visibility in search engines and different vocabularies like DC, FOAF can be used with RDFa.

- **Requirements to implement RDFa**

Interviewee 5 said that if the website is written in HTML then the code should be converted into XHTML1.1.

- **Technical steps of implementing RDFa**
According to Interviewee 5 the following technical steps involved in using of RDFa in structuring of websites: First step is converting the code into XHTML 1.1 because RDFa does not support HTML code. HTML is not extensible. Second step is adding the DOCTYPE XHTML+RDFa1.0 and if it is required to used the vocabularies the DOCTYPE XHTML+RDFa1.1 will use. Third step is using the RDFa’s attributes according to the needs and which vocabulary will use. Forth step is converting the URIs into CURIES. Fifth step is the root element should be HTML.

- **Vocabulary used.**

Interviewee 5 used FOAF and DC vocabularies and he is planning to use the GoodRelations in structuring the websites.

- **Barriers of RDFa**

The big barrier RDFa is that it can not be used in HTML and most of the websites have written in html, the other barriers are: XHTML cleaning tool can not be used because it affect the RDF triples and URIs should be converted into CURIEs.

- **How semantic markup enhances web visibility**

According to interviewee 5 RDFa makes the code machine readable. It is easy for the crawlers or spiders in SEO process to find the Web sites those are structured in RDFa. This leads to increase the web visibility of web sites in search engines.

- **Evaluation tools that used**

Interview 5 evaluated the website visibility in search engines by using different tools but commonly he used Google Analytics.

- **Benefits of RDFa**

According to Interviewee 5 there are many benefits of RDFa like easy to use, make the code machine readable by adding the semantics, different vocabularies can be used with RDFa and the websites structured with RDFa are more visible in search engines.

- **Difficulties in implementing RDFa**
Interviewee 5 did not find any difficulty in using the RDFa.

**Interviewee 6:**

The sixth interviewee works as web developer in a prominent software house in Pakistan.

- **Motivation behind using RDFa**

Interviewee 6 has implemented RDFa in structuring the websites. According to him/her his company has to pay a lot of money to make high ranking of their website in search engines. They were trying to find the solution to make the website’s ranking without paying much money, so he implemented the RDFa in company’s website, as a result company’s website has gotten high ranking in search engine.

- **Requirements to implement RDFa**

Interviewee 6 explained that there are few requirements to implement the RDFa and those are: Code should be written in XHTML1.1, DOCTYPE declaration, selection of vocabularies.

- **Technical steps of implementing RDFa**

According to Interviewee 6 technical steps that involved in using of RDFa are: the root element should be HTML, code should be in XHTML1.1 because HTML is not extensible, add the DOCTYPE XHTML+RDFa1.0 in the header, converting the URIs into CURIES because in URIs it needs to write the complete URIs but in CURIES main part of the URIs assigned to token and during the processing CURIEs converted to URIs.

- **Vocabularies**

According to interviewee 6, he has used FOAF and GoodRelations vocabularies.

- **Barriers of using RDFa**

Interviewee 6 has said that there are few barrier of using RDFa and those are code should be in XHTML 1.1, XHTML cleaning tool cannot be used because it effect the RDF triples and URIs should be converted into CURIEs.
- **How semantic markup enhances web visibility.**

Web sites structured in RDFa are machine readable so it is easy for the browsers to understand the contents of the web sites. As a result, the web visibility of web sites enhance in search engines.

- **How evaluate web visibility.**

Interview 6 has evaluated the website visibility by using Google Analytics.

- **Benefits of RDFa**

According to him/her RDFa is very useful semantic markup technique, it is easy to implement, with the help of RDFa RDF triples can implement in the websites, it make the cod reuse able, different vocabularies can be used with RDFa and implementation of RDFa save the money

- **Difficulties in implementing RDFa**

According to interviewee 6 sometime it is hard job to convert the HTML code into XHTML1.1 to implement RDFa.