

Systematic implementation of Schema.org on Product Offers



Farrukh Butt (77-BSCS-12)

Ahmad Saleem (113-BSCS-12)

Session 2012-2016

DEPARTMENT OF COMPUTER SCIENCE

GC UNIVERSITY LAHORE

Systematic implementation of Schema.org on Product Offers

Submitted to GC University Lahore
in partial fulfillment of the requirements
for the award of degree of
Bachelor's
IN
Computer Science

By

NAME: Farrukh Butt

SESSION: 2012-2016

Roll No: 77-BSBS-12

NAME: Ahmad Saleem

SESSION: 2012-2016

Roll No: 113-BSBS-12

DEPARTMENT OF COMPUTER SCIENCE

GC UNIVERSITY LAHORE

DECLARATION

We **Ahmad Saleem** Roll No.**113-BSCS-12** and **Farrukh Butt** Roll No.**77-BSCS-12** students of **Bachelor of Science** in the subject **of Computer Science** session **2012-2016**, hereby declare that the matter printed in the thesis titled **Systematic implementation of Schema.org on Product Offers** is my own work and has not been printed, published and submitted as research work, thesis or publication in any form in any University, Research Institution etc in Pakistan or abroad.

Dated: _____

Ahmad Saleem

Farrukh Butt

RESEARCH COMPLETION CERTIFICATE

Certified that the research work contained in this thesis titled **Systematic Implementation of Scheme.Org on Product Offer** has been carried out and completed by **Mr. Ahmad Saleem Roll No.113-BSCS-12 and Farrukh Butt Roll No.77-BSCS-12,** **Under my supervision.**

Date

Supervisor Signature
Mr. Imran Rafique

Submitted Through

Prof. Prof. Dr. Asad Raza Kazmi

Chairperson

Department of _____

GC University Lahore.

Controller of Examinations

GC University Lahore.

ACKNOWLEDGEMENT AND DEDICATION

We would like to Thanks Allah Almighty who gave us capability and patience to complete our project. We would specially like to thank Mr. Imran Rafique who guided us and dedicated his time and efforts for this project. We are also grateful to all the teachers of Department of Computer Science who not only taught us the curriculum courses but also enlightened us with moral values.

We would also like to remember all the students of 2012 batch who were the part of our 4 year splendid journey.

We dedicate all our work to our Parents!

Their support, courage and love

Have made it possible for us.

ABSTRACT

Systematic implementation of schema.org on product offer is a serious effort towards the automatic implementation of schema tags on raw data.

Schema.org tags are a joint venture of google and bing in the world of semantic web. Schema tags are implemented in HTML and they help search crawlers to understand the content of a web page and as a result of it that webpage gets a better search ranking.

Implementing schema tags manually is very error prone and tedious work. The goal of our project is to implement them automatically. Our system take help of different analysis techniques to filter out information like keyword analysis and Natural language processing.

For the implementation of our system we have developed a responsive website that gets data from a web service, saves it into its database and then analyze that information for the implementation of Schema tags.

Contents

| | |
|--|-----------|
| GC UNIVERSITY LAHORE..... | 1 |
| INTRODUCITON | 1 |
| SCHEMA.ORG TAGS | 1 |
| PROBLEM STATEMENT..... | 1 |
| PROPOSED SOLUTION..... | 2 |
| <i>Capturing Data</i> | 2 |
| <i>Filtering Mechanism</i> | 3 |
| <i>Saving filtered information in database</i> | 3 |
| <i>Website</i> | 3 |
| LITERATURE REVIEW | 4 |
| GOOGLE SEARCH..... | 4 |
| GOOGLE PENGUIN | 4 |
| GOOGLE HUMMING BIRD..... | 4 |
| GOOGLE RANKBRAIN | 5 |
| GOOGLE PANDA ALGORITHM | 6 |
| <i>Thin Content</i> | 6 |
| <i>Aggregated content</i> | 6 |
| <i>Duplicate Content</i> | 6 |
| <i>Irrelevant Content</i> | 6 |
| UNDERSTANDING THE CONTENT | 7 |
| <i>By looking into the structure</i> | 7 |
| <i>By looking into Meta tags</i> | 7 |
| <i>Google NLP or machine learning</i> | 7 |
| WHY SCHEMA.ORG..... | 7 |
| <i>GoodRelations Code</i> | 8 |
| <i>Schema.Org Code</i> | 9 |
| SCHEMA IMPLEMENTATION | 11 |
| <i>Form Based Implementation</i> | 12 |
| <i>RDFa Content Editor by Mr. Ali Khalili</i> | 12 |
| NEED OF DEVELOPMENT OF OUR NATIVE SOLUTION..... | 13 |
| REQUIREMENT SPECIFICATION..... | 14 |
| FUNCTIONAL REQUIREMENTS..... | 14 |
| <i>Capture Information Online</i> | 14 |
| <i>Saving Captured information</i> | 14 |
| <i>Website</i> | 14 |

| | |
|---|-----------|
| <i>Admin performing its roles</i> | 14 |
| <i>User performing its roles</i> | 15 |
| NONFUNCTIONAL REQUIREMENT | 16 |
| <i>Performance</i> | 16 |
| <i>Security</i> | 16 |
| <i>Integrity</i> | 16 |
| PROJECT DESIGN | 17 |
| <i>Use Case Diagram</i> | 17 |
| <i>Use Case Scenarios</i> | 20 |
| DESIGN DESCRIPTION | 29 |
| <i>System Sequence Diagram</i> | 29 |
| <i>Sequence Diagrams</i> | 31 |
| <i>Entity Relationship Diagram</i> | 35 |
| <i>Data Tables</i> | 39 |
| <i>Class Diagram</i> | 43 |
| <i>Activity Diagram</i> | 44 |
| IMPLEMENTATION | 52 |
| DEVELOPMENT STAGES | 52 |
| <i>Research</i> | 52 |
| <i>Planning</i> | 52 |
| <i>Design</i> | 52 |
| <i>Development:</i> | 52 |
| <i>Testing</i> | 53 |
| USER INTERFACE | 53 |
| <i>Main Screen</i> | 53 |
| <i>Advertiser screen</i> | 54 |
| <i>Product Screen</i> | 55 |
| <i>Offer Screen</i> | 55 |
| <i>Login Screen</i> | 56 |
| <i>Register new admin</i> | 56 |
| <i>Main Admin Screen</i> | 57 |
| <i>Admin Advertiser screen</i> | 57 |
| <i>Admin Advertiser Edit Screen</i> | 58 |
| <i>Admin Advertiser Detail</i> | 58 |
| <i>Admin Product screen</i> | 59 |
| <i>Admin Product Edit Screen</i> | 60 |
| <i>Admin Product Detail Screen</i> | 60 |
| <i>Admin offers Screen</i> | 61 |
| <i>Admin offer detail</i> | 61 |

| | |
|--|-----------|
| <i>Offer Edit</i> | 62 |
| <i>Offer Add</i> | 62 |
| <i>NLP NER Analysis</i> | 63 |
| <i>Scrapping Screen</i> | 63 |
| <i>Keyword Cloud</i> | 64 |
| EVALUATION | 65 |
| TEST CASES | 65 |
| <i>Admin Login Test Case 1</i> | 65 |
| <i>Admin Login Test Case 2</i> | 66 |
| <i>Admin Rights Test Case</i> | 66 |
| <i>Admin editing advertisers test case</i> | 67 |
| <i>Admin editing products test case</i> | 68 |
| <i>Admin editing offers</i> | 68 |
| <i>Admin performing keyword analysis over offers test case</i> | 69 |
| <i>Admin performing NER/NLP analysis over offers test case</i> | 70 |
| <i>Admin adding offers test case</i> | 70 |
| <i>Approved offers information test case</i> | 71 |
| <i>Admin scrapping website test case</i> | 72 |
| <i>Admin getting top keywords test case</i> | 73 |
| <i>User searching products test case</i> | 73 |
| <i>Particular category offers test case</i> | 74 |
| <i>Particular advisers offer test case</i> | 74 |
| RESULTS | 75 |
| CONCLUSIONS & FUTURE WORK | 76 |
| REFERENCES | 77 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1: Wife of President..... | 5 |
| Figure 2: Form Based Implementation[1]..... | 12 |
| Figure 14: Use Case Diagram | 17 |
| Figure 15: User Use Case | 18 |
| Figure 16: Admin Use Case | 18 |
| Figure 17: Manage Advertisers..... | 19 |
| Figure 18: Manage Offers | 19 |
| Figure 19: Manage Products | 19 |
| Figure 3: SSD User | 29 |
| Figure 4: SSD Admin..... | 30 |
| Figure 5: Add Offers Sequence Diagram..... | 31 |
| Figure 6: Edit Offers Sequence Diagram..... | 31 |
| Figure 7: Edit Advertiser Sequence Diagram | 32 |
| Figure 8: Edit Offers Sequence Diagram..... | 32 |
| Figure 9: View Offers Sequence Diagram..... | 33 |
| Figure 10: View Products Sequence Diagram | 33 |
| Figure 11: View Advertisers Sequence Diagram..... | 34 |
| Figure 12: ERD | 35 |
| Figure 13: Class Diagram | 43 |
| Figure 20: Activity Diagram..... | 44 |
| Figure 21: Approve Offers..... | 45 |
| Figure 22: Advertiser Detail | 46 |
| Figure 23: Offer Detail..... | 47 |
| Figure 24: Product Detail..... | 48 |
| Figure 25: Advertiser Edit | 49 |
| Figure 26: Offer Edit..... | 50 |
| Figure 27: Edit Product..... | 51 |
| Figure 28: Home Page | 54 |
| Figure 29: Advertiser Screen | 55 |
| Figure 30: Product Screen..... | 55 |
| Figure 31: Offer Screen | 56 |
| Figure 32: Login Screen..... | 56 |
| Figure 33: Register Screen..... | 57 |
| Figure 34: Admin Home Page | 57 |
| Figure 35: Advertiser Detail Screen | 58 |
| Figure 36: Advertiser Edit Screen..... | 58 |
| Figure 37: Advertiser Details Screen..... | 59 |
| Figure 38: Products Screen | 59 |
| Figure 39: Products Edit Screen | 60 |
| Figure 40: Products Detail Screen | 60 |
| Figure 41: Offers Screen | 61 |
| Figure 42: Offer Detail Screen..... | 61 |
| Figure 43: Offer Edit Screen..... | 62 |
| Figure 44: Add Offers Screen | 62 |
| Figure 45: NER Analysis | 63 |
| Figure 46: Scraper..... | 63 |
| Figure 47: Key Cloud..... | 64 |

LIST OF TABLES

| | |
|---|----|
| Table 12: Actor Admin | 20 |
| Table 13: Actor User..... | 20 |
| Table 14: Use Case Edit Advertiser | 20 |
| Table 15: Use Case Detail of Advertiser | 21 |
| Table 16: Use Case Delete Advertiser | 22 |
| Table 17: Use Case Edit Product | 22 |
| Table 18: Use Case Detail Product | 23 |
| Table 19: Use Case Delete Product | 23 |
| Table 20: Use Case Edit Offers | 24 |
| Table 21: Use Case Detail Offers | 24 |
| Table 22: Use Case Approve Offers | 25 |
| Table 23: Use Case Delete Offers..... | 26 |
| Table 24: Use Case Analysis Cloud..... | 26 |
| Table 25: Use Case Scrapper | 27 |
| Table 26: Use Case View Advertisers | 27 |
| Table 27: Use Case View Products..... | 27 |
| Table 28: Use Case View Offers | 28 |
| Table 29: Use Case Search | 28 |
| Table 1: Data Table Action..... | 39 |
| Table 2: Data Table Advertiser_LinkType | 39 |
| Table 3:Data Table Advertiser..... | 40 |
| Table 4: Data Table Category | 40 |
| Table 5: Data Table Commission | 40 |
| Table 6: Data Table Language | 41 |
| Table 7: Data Table LinkType | 41 |
| Table 8: Offer..... | 41 |
| Table 9:Data Table OfferDetail | 42 |
| Table 10: Data Table Promotion..... | 42 |
| Table 11: Data Table Product | 42 |
| Table 30: Test Case Login | 65 |
| Table 31:Test Case Login 2 | 66 |
| Table 32:Test Case Admin Rights | 66 |
| Table 33:Test Case Edit Advertiser | 67 |
| Table 34: Test Case Edit Product..... | 68 |
| Table 35: Test Case Edit Offers..... | 68 |
| Table 36:Test Case Key Analysis | 69 |
| Table 37:Test Case NER..... | 70 |
| Table 38:Test Case Admin Adding Offers | 70 |
| Table 39: Test Case Approved Offers | 71 |
| Table 40:Test Case Scrapping | 72 |
| Table 41:Test Case Top Key Words..... | 73 |
| Table 42: Searching Products | 73 |
| Table 43:Test Case Category Offer | 74 |
| Table 44: Particular Advertiser..... | 74 |

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

INTRODUCITON

Schema.Org tags

Schema.Org is an initiative by Google and Bing to create a common vocabulary for structured data markup on web pages[7] . The purpose of schema.org is that webmasters could markup their own content with metadata so that web crawlers could understand the content and rank it. Schema.Org vocabulary could be implemented while using Microdata, RDFa or JSON-LD formats.

Implementing schema.org in your site is an important factor if you want to increase the search ranking of your site. As google Panda algorithm totally ranks the content upon its quality. According to a latest survey only 1% of sites are currently implementing schema tags but there is always at least one site on first page of google search that has implemented schema tags.

Problem Statement

Google introduces its panda algorithm in 2011 which was intended to filter low quality sites upon the basis of their content. Upon the introduction of this ranking algorithm many sites saw the drop in their ranking and some of them were genuine sites which were wrongly hit.

If you want to save your site being negatively ranked from panda algorithm then it is important that you have a good quality content upon your site moreover the web crawlers should be able to understand that content so that they could rank it. In case if crawler fails to understand your content then it would result in down ranking of your site.

To help crawlers and assist web masters Google and Bing introduced Schema.Org vocabulary on 2nd June 2011 just after 4 months of the introduction of panda algorithm. As described above Schema.org is a vocabulary so that webmasters could markup the content of their site.

Websites now a days contains hundreds of thousands of records and are updated upon daily basis. In such a scenario it is very hard to implement schema tags manually. It is error prone and can cause problems and drop in ranking instead of increase in search ranking. Moreover it is also very tedious job and requires a full time skillful person to get the job done.

Our project is first step towards the automation of implementation of schema tags. The main challenge for us was not to markup the data but to extract some meaningful information out of raw data. Implementation of Schema tags was only possible after the filtration of meaningful information and then bringing it into structured form.

Initially we tried to use Natural Language Processing to filter out the information but our experimentation with NLP[9] didn't bring fruitful results as the content we were receiving was not in proper grammatical or semantic structure. For this we had to use other techniques like keyword analysis for information extraction.

Proposed Solution

Our proposed solution comprised of following points

- 1) Capturing of data.
- 2) Construction of a mechanism that filters out information from data.
- 3) Save filtered out information in a database in a structured form.
- 4) Creation of a website that take information from that database and displays it.

Capturing Data

We are working upon implementation of schema tags upon products and offers so we naturally needs its real time data in real time environment for this purpose we are using a web api that delivers us information and we are initially dumping it in a database

Filtering Mechanism

Once data is in database it is needed to be filtered and to be saved back in database. For this purpose we are using 2 different filtering techniques .The first one is NLP and the second one is keyword search or pattern analysis. We would be discussing these techniques in detail further in this document.

Saving filtered information in database

Once we filter out information with the help of keyword analysis and NLP then we save it back to database for future use.

Website

Our website is core of our project and demonstrates the working of our mechanism in detail. It is a comprehensive website that has a user and admin side and allow user and admin to perform actions of their interest. Admin can edit view and analyze information whereas user can see products and offers and search down information of his interest.

LITERATURE REVIEW

Google Search

Google uses several different algorithms to rank a website and to process user query. The four main algorithms are listed below

- 1) Google Penguin
- 2) Google Humming Bird
- 3) Google Rank Brain
- 4) Google Panda Algorithm

Google Penguin

Google also use to decide a rank of a website upon the basis of its backlinks. Backlinks are links to your website from other websites. It is quite natural that if a site has a good content over it then other sites would link themselves to it and will give its reference in their content. So google considers each backlink as a vote to that site and increases its ranking. However people exploited this policy of google and created fake links to their site by techniques like link farms, links in fake websites, comments and self-written articles. To avoid being tricked and saving users from bad search results google introduced its penguin algorithm back in 2012 [3]. The main purpose of the penguin algorithm is to search for the links of website upon the internet and checking their validity. In case if it find any black hat way of increasing links to your site then it down score your site.

Google Humming Bird

Google humming bird is an algorithm to understand user query and return results that best matches (Google Algorithms, n.d.) for instance if you write this query upon google “Wife of president of America” then it will return you Michelle Obama. Note that you neither mentioned her name or the name of president but google returned you the right result which is only possible

because of google humming bird algorithm. One such demonstration is pasted below.

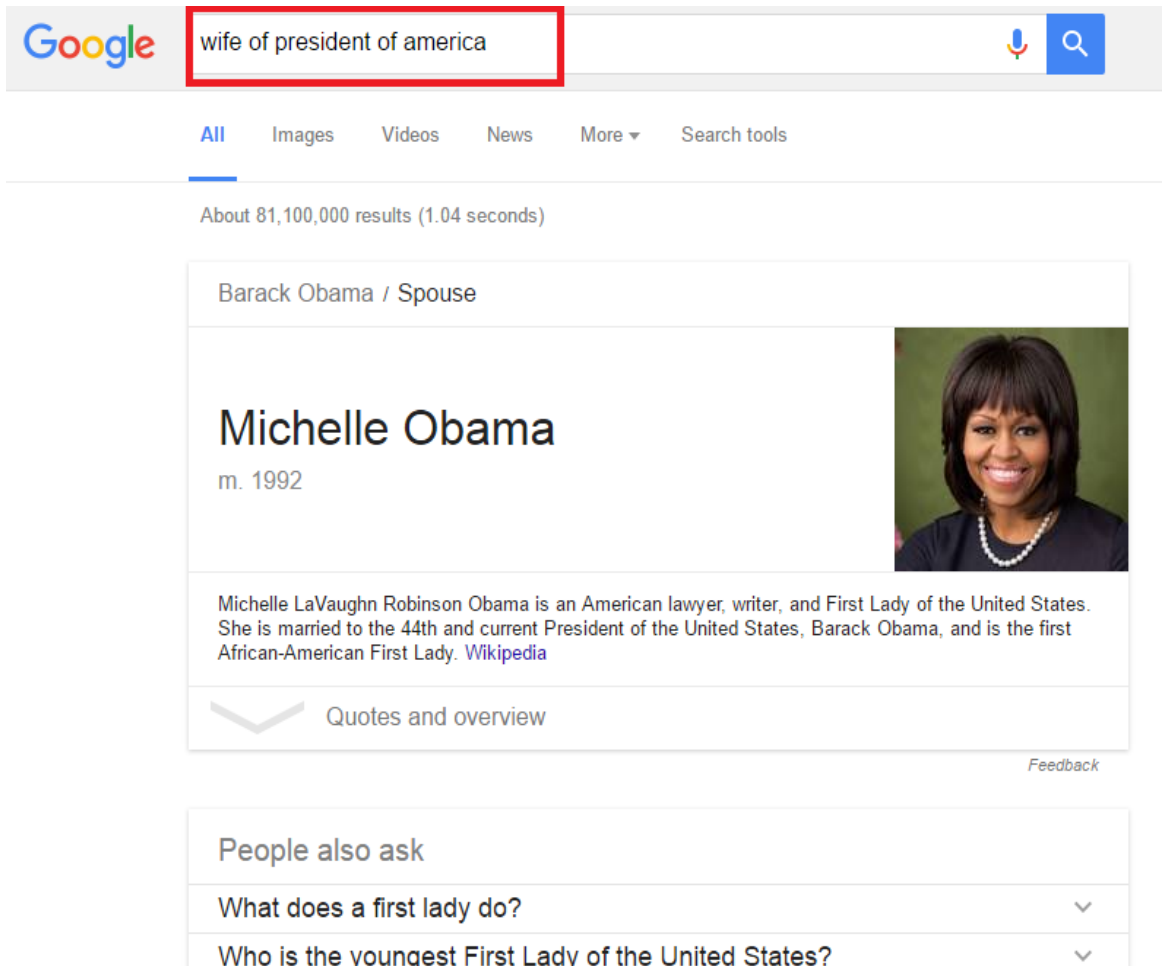


Figure 1: Wife of President

Google RankBrain

Google RankBrain is based upon Artificial Intelligence and Machine learning[6] currently it is being used as a component of penguin algorithm to understand queries but google has announced that it will also be used to decide the rank of a page.

Google Panda Algorithm

Google Panda Algorithm works upon deciding the quality of content on a webpage and is center of our research and project[4]. It was introduced in 2011 with the intention of filtering out low quality content and to provide user a good experience. As it is center of our interest so we will look it into detail that how it works.

Google mentions 23 points upon the basis of which panda works or ranks the quality of a page which can be viewed in this blog post [5].

However summarizing up things we can summarize it down to five major points

Thin Content

By thin content it means that your page contain very few or little information about the topic. If your content is thin then neither it will satisfy user nor will it help search crawlers to understand or rank your content

Aggregated content

Aggregated content means that you have just aggregated the contents from different sites and displayed it upon your site. In this case your content is neither unique nor original so google panda also down ranks page with aggregated content

Duplicate Content

If content on your page is too similar or copied from other sites then it is considered as duplicate content and is also negatively ranked by google

Irrelevant Content

By irrelevant content it means that your site doesn't provide the content that it promise to e.g. the content and title are different or the content discusses the actual topic for a very short time and majority of it consist upon things that are not directly related to topic under consideration.

Understanding the Content

The first step in google panda algorithm is to understand the content. This the part where schema.org comes into action. Google Panda understands the content through various techniques.

By looking into the structure

Google understands the content by looking into the structure of document. By ranking h1 heading as the most important par and h2 as less important and same goes down the tags

By looking into Meta tags

Meta tags provide crawlers an overall look or the general information about the content of page. Meta tags can provide an overview of page but can't provide a deep dive into data.

Google NLP or machine learning

Google also tries to understand the content by using NLP and machine learning algorithms.

Beside all this there is still a great possibility that google may not understand your content or may develop a wrong sense of your content. To save yourself from such a scenario it is important to tag your content with schema tags. It will not only help google to understand your content but it will also increase your SEO.

Why Schema.org

The schema.org vocabulary was launched in 2011 however the history of semantic web is very old and a lot of libraries do existed before schema.org that provide a much larger and specialized vocabulary to tag your site. One such vocabulary in which we looked into was GoodRealtions by purl.org. [2] It is business oriented and offers a wide range of tags to mark your content. So the question arises Why to use Schema.org tags instead of others? The simple one liner answer is "Because Google, Yahoo and Bing say so". As I described earlier that Schema.org is a joint project by major search engines and has a strong backing and offers a promised growth. If we use any other vocabulary than schema then our whole point of using it goes invalidated

because they are no longer supported by search engines and a crawler won't understand them. In case of GoodRealtions it has been merged into Schema.Org and all its vocabulary could be accessed through Schema.Org namespace. To further prove our point we marked same content with GoodRelations and Schema.Org to show that Schema is now as rich as goodrelations. Both of the code snippets are pasted below.

GoodRelations Code

```
<html>
```

```
<body>
```

```
<div itemscope itemtype="http://purl.org/goodrelations/v1#Offering">
```

```
<div itemprop="name">
```

Hepp Personal SCSI Controller Card

```
</div>
```

```
<div itemprop="description">
```

The Hepp Personal SCSI is a 16-bit

add-on card that allows attaching up to seven SCSI devices to your computer.

```
</div>
```

```
<link itemprop="hasBusinessFunction"
```

```
href="http://purl.org/goodrelations/v1#Sell" />
```

```
<div itemscope itemprop="hasPriceSpecification"
```

```
itemtype="http://purl.org/goodrelations/v1#UnitPriceSpecification">
```

Price:

```
<meta itemprop="hasCurrency" content="USD">
```

\$

```
<span itemprop="hasCurrencyValue">
```

99.99

```
</span>
```

```
<time itemprop="validThrough" datetime="2012-11-30T23:59:59Z"></time>
```

</div>

Condition: <div itemprop="condition">used</div>

EAN/UPC: 1234567890123

MPN: PSCSI

Article No. 123-456

Availability: <span itemscope itemprop="hasInventoryLevel"

itemtype="http://purl.org/goodrelations/v1#QuantitativeValue">

<meta property="hasMinValue" content="1">In-stock

</div>

</body>

</html>

Schema.Org Code

<html>

<body>

<div itemscope itemtype="https://schema.org/Offer">

<div itemprop="name">

Hepp Personal SCSI Controller Card

</div>

<div itemprop="description">

The Hepp Personal SCSI is a 16-bit

add-on card that allows attaching up to seven SCSI devices to your computer.

</div>

<link itemprop="businessFunction"
href="http://purl.org/goodrelations/v1#Sell" />

<div itemscope itemprop="priceSpecification"
itemtype="https://schema.org/PriceSpecification">

Price:

<meta itemprop="priceCurrency" content="USD">

\$

99.99

<time itemprop="validThrough" datetime="2012-11-30T23:59:59Z"></time>

</div>

Condition:

<div itemprop="condition" href="http://schema.org/UsedCondition">

Used

</div>

EAN/UPC:

1234567890123

MPN:

PSCSI

Article No.

123-456

Availability:

```
<span itemscope itemprop="inventoryLevel"
      itemtype="https://schema.org/QuantitativeValue">
  <meta property="minValue" content="1">
```

In-stock

```
</span>
```

```
</div>
```

```
</body>
```

```
</html>
```

From the above example it is clearly evident that Schema.Org offers you as rich as vocabulary as GoodRelations.

Schema Implementation

After all this research we find out that schema is something that is being neglected by many developers whereas it should be the part of every website because it increases the understandability of your content to crawlers and in return you get a better search ranking and good user feedback.

The major reason of not implementing schema is the overhead of implementing it. People don't usually have idea how it works and the manual way of implementing is very tedious and error prone work. To overcome such shortcomings it is necessary to define some automatic or semi-automatic way of implementing schema.org

After initial work we started looking for models or system that use to implement schema.org automatically or semi automatically, we came two such systems

- 1) Form Based Implementation
- 2) RDFa Content Editor by Mr. Ali Khalili

Form Based Implementation

One example of implementing schema on your information that we came across is through manually filling a form and resulted schema would be generated. One such demonstration is pasted below

The screenshot shows the 'Microdata Generator Using Schema.org + JSON-LD' website. The sidebar on the left lists various business types: Animal Shelter, Automotive, Child Care, Dry Cleaning, Emergency Service, Employment Agency, Entertainment, Financial Service, Restaurant / Food Service, Post Office, Health and Beauty, Home and Construction, Internet Cafe, Legal, Lodging, and Media Broadcasting. The main area is titled 'Step 1: Choose a Business Type' and features a 'Generate Code' button. Below this is a form with fields for Business Name, Website URL, Email, Phone, Street, City, State, Zip, Rating, and Review Count. The 'JSON-LD Code' box on the right displays the generated schema code, which includes a script tag and a JSON-LD object for a business named 'my business'.

```
<script type="application/ld+json"> {  
  "@context": "http://schema.org",  
  "@type": "Step 1: Choose a Business Type",  
  "name": "my business",  
  "email": "abc@example.com",  
  "telephone": "123",  
  "priceRange": "4"  
}</script>
```

Figure 2: Form Based Implementation[1]
(Form based schema implementation, n.d.)

These kind of form filling techniques are good for learning purposes or to get an idea of how information would be presented in hierarchy and then manually replicate the information in your page. But implementing these models to directly implement schema in your site is not practical.

RDFa Content Editor by Mr. Ali Khalili

One great effort to implement schema systematically is done by Mr. Ali khalili. In his paper (The RDFa Content Editor - From WYSIWYG to WYSIWYM.)[8] that received best paper award he explains that how systematic implementation of schema.org could be achieved. Later on he developed a system based upon his research and distributed different versions of it. This system is NLP based system. In his system Mr. Ali khalili has merged different NLP libraries and analyze content with the help of them and generate suggestions based upon analysis and also provide a form and a systematic way of implementing schema on a raw text. We took a look at their paper and saw their demonstration video but unluckily we were unable to get a working

model of their system. Moreover as this is system was NLP based so in our case we had limitations with nlp.

Need of development of our native solution

After taking a look at both we concluded that form based solution is not a practical approach to be implemented because usually information is not structured and has to be refined moreover a user may would like to present information other than table format. Mr Ali Khalili kind of model was much suitable in our case but unfortunately we were unable to find out a working model of system so we were not able to verify its accuracy independently moreover this system was NLP based but the data that we have is related to products and offers and wasn't in proper grammatical or semantic structure because of it only NLP wasn't much affective in filtering out the information. The data that we have regarding different products and offers lack grammatical structure but it has different patterns in it so by identifying these patterns we can also filter out meaningful information. Additionally we needed a system that not only filters out information from a content but it should also be able to capture data online and save it in a database and present it in front of user and also enable admin to edit and modify contents.

Requirement Specification

Functional Requirements

We have identified following functional requirements

Capture Information Online

This component require us to design a way to capture data of products and offers online either via scrapping or through some api.

Saving Captured information

Once the information is data is captured then it must be converted into tabular form and stored into database

Website

- Accessible to both user and admin
- Responsive
- Dynamic

Admin performing its roles

- Admin should able to edit/delete products
- Admin should be able to edit/delete advertisers
- Admin should be able to edit/delete offers
- Admin should be able to approve offers
- Admin should be able to perform Keyword analysis over offers
- Admin should be able to perform NLP analysis over offers
- Admin should be able to navigate to products and offers through advertisers
- Admin should be able to search a particular advertiser, Product or offer
- Admin should be able to scrap a site of its choice
- Admin should be able to see top keywords in database

User performing its roles

- User should be able to look newest offers
- User should be able to look top offers
- User should be able to look top Products
- User should be able to look products in Alphabetical order
- User should be able to look top Advertisers
- User should be able to look Advertisers in alphabetical order
- User should be able to look products of particular advertiser
- User should be able to look offers of a particular advertiser
- User should be able to look products by category
- User should be able to look advertisers by category
- User should be able to look offers by category
- User should be able to search offers
- User should be able to search products
- User should be able to search advertisers

Nonfunctional Requirement

Performance

We need to perform a lot of processing and amount of data is large so efficient coding and design is necessary to provide a better performance.

Security

System should be secured against threats like sql injection and other potential hacks.

Integrity

The information that we present or the analysis we perform should be accurate to maximum extent and rate of errors should be brought down to a minimal level.

PROJECT DESIGN

Use Case Diagram

Complete System Use case

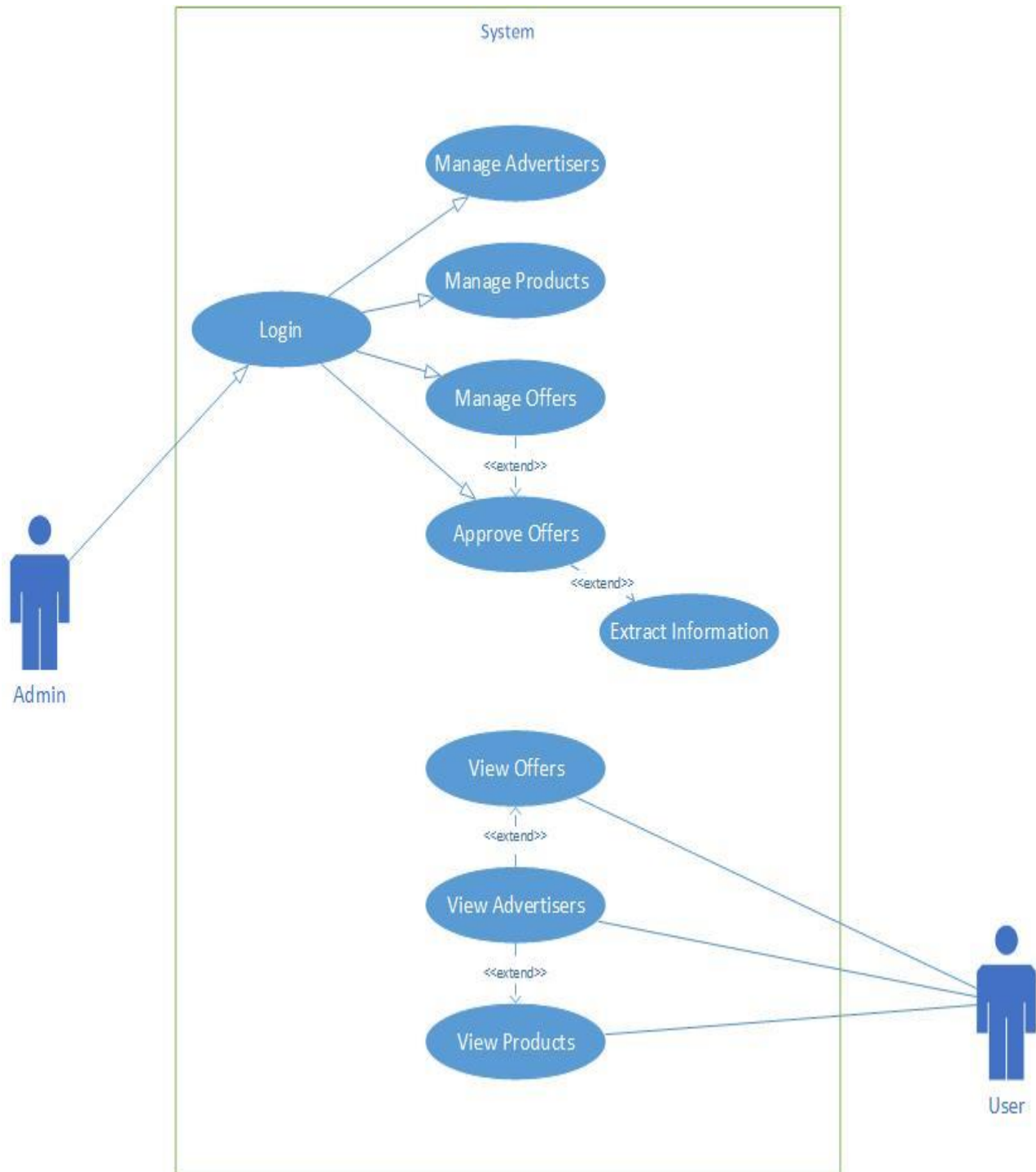


Figure 3: Use Case Diagram

User Use Case

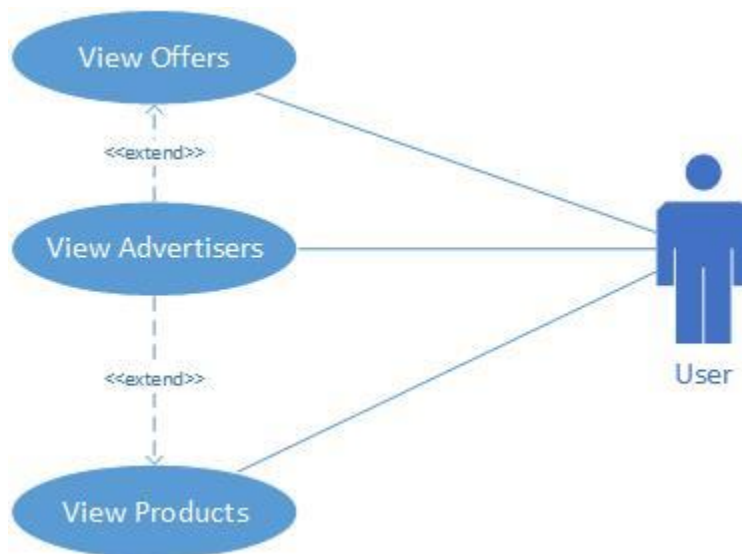


Figure 4: User Use Case

Admin Use Cases

Approve offers

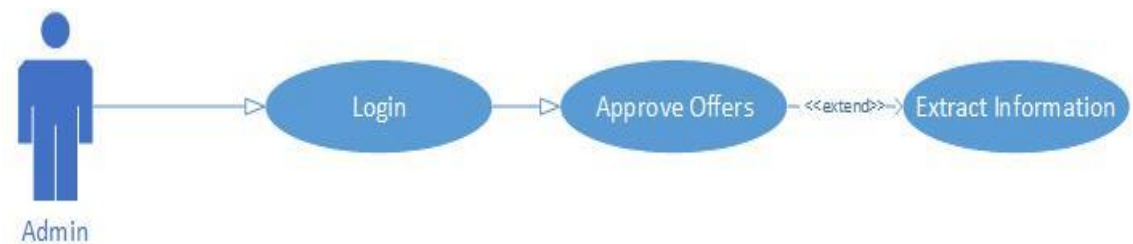


Figure 5: Admin Use Case

Manage Advertisers

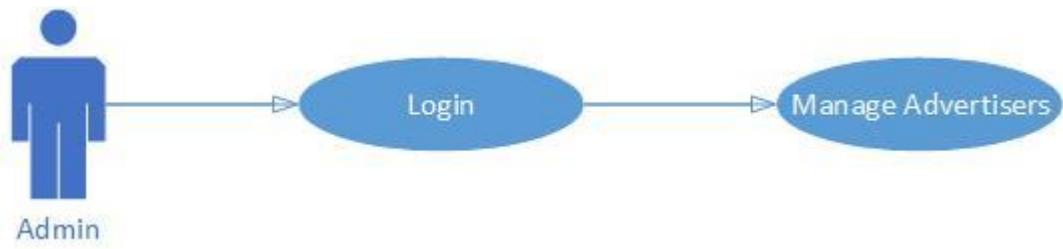


Figure 6: Manage Advertisers

Manage Offers

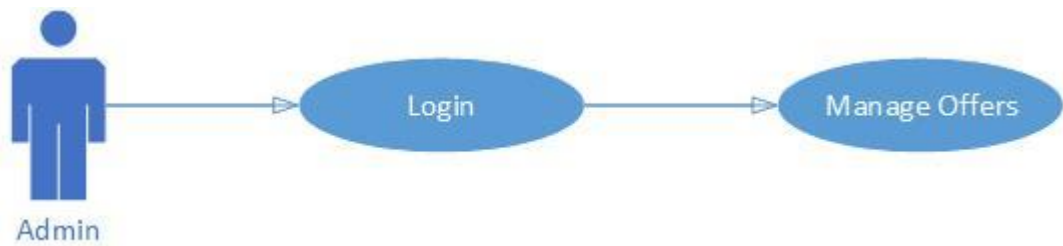


Figure 7: Manage Offers

Manage Products

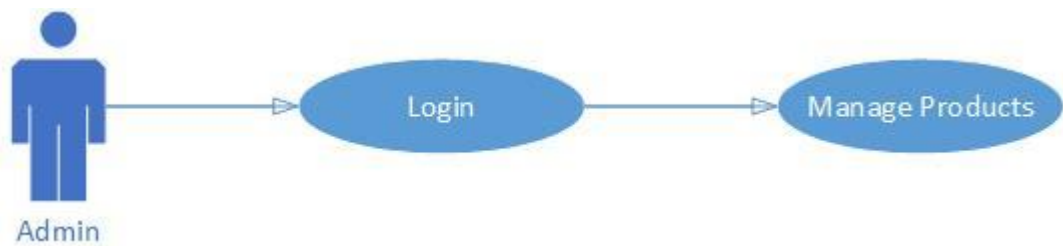


Figure 8: Manage Products

Use Case Scenarios

Specification of actors

The following actors are defined so far in the analysis phase of system development process.

Admin

Table 1: Actor Admin

| Admin | |
|-------------|--|
| Element | Details |
| Description | This actor will handle all the admin side of the web site and manage the operations. |
| Examples | Manage the Advertisers, Products and Offers and Approve offers. |

User

Table 2: Actor User

| User | |
|-------------|--|
| Element | Details |
| Description | User is the simple use of the web site which will view advertisers, products and offers. |
| Examples | View latest offers in arranged form. |

Specification of Use Case

The figure shows the Use Case Specification that might be changed during later stages in the development process.

Edit Advertiser

Table 3: Use Case Edit Advertiser

| Edit Advertiser | |
|-----------------|--|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to edit the information of the Advertiser |
| Pre-Conditions | Advertiser ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |

| Edit Advertiser | |
|---------------------|--|
| Element | Details |
| Post Conditions | Advertiser is modified in the system, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Advertisers from Admin Main Page2. All advertiser are appeared in the web page3. Admin select EDIT Advertiser option of the specific advertiser from the web page4. Admin edit the Advertiser according to its need and click save button5. System updates |
| Alternative courses | Not all mandatory data fields are filled |

Detail of Advertiser

Table 4: Use Case Detail of Advertiser

| Detail Advertiser | |
|---------------------|--|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to view Detail information of the Advertiser and view Offers and Products of specific Advertiser or Advertisers Actions and Commissions |
| Pre-Conditions | Advertiser ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Advertiser details is viewed, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Advertisers from Admin Main Page2. All advertiser are appeared in the web page3. Admin select Details Advertiser option of the specific advertiser from the web page4. Admin can now select Offers and Products of selected Advertiser or Actions and Commissions |
| Alternative courses | Admin select Back To List Option to View All Advertisers |

Delete Advertiser

Table 5: Use Case Delete Advertiser

| Delete Advertiser | |
|---------------------|---|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to delete the Advertiser |
| Pre-Conditions | Advertiser ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Advertiser is deleted from the system, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Advertisers from Admin Main Page2. All advertiser are appeared in the web page3. Admin select Delete Advertiser option of the specific advertiser from the web page4. Admin conform the deletion of the advertiser5. System updates |
| Alternative courses | Admin does not conform the deletion |

Edit Products

Table 6: Use Case Edit Product

| Edit Products | |
|-----------------|---|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to edit the information of the Products |
| Pre-Conditions | Products ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Products is modified in the system, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Products from Admin Main Page2. All Products are appeared in the web page3. Admin select EDIT Products option of the specific advertiser from the web page4. Admin edit the Products according to its need and click save button5. System updates |

| Edit Products | |
|---------------------|--|
| Element | Details |
| Alternative courses | Not all mandatory data fields are filled |

Detail Product

Table 7: Use Case Detail Product

| Detail Product | |
|---------------------|---|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to view Detail information of the Product |
| Pre-Conditions | Product ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Product details Viewed, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Produits from Admin Main Page2. All Product are appeared in the web page3. Admin select Details Product option of the specific Product from the web page4. Admin can now select View Detailed information about products |
| Alternative courses | Admin select Back To List Option to View All Products |

Delete Products

Table 8: Use Case Delete Product

| Delete Products | |
|-----------------|--|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to delete the Products |
| Pre-Conditions | Products ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Products is deleted from the system, the user is logged into the system, and the system menu is displayed. |

| Delete Products | |
|---------------------|--|
| Element | Details |
| Normal course | <ol style="list-style-type: none">1. Admin select Products from Admin Main Page2. All Products are appeared in the web page3. Admin select Delete Products option of the specific advertiser from the web page4. Admin conform the deletion of the Products5. System updates |
| Alternative courses | Admin does not conform the deletion |

Edit Offers

Table 9: Use Case Edit Offers

| Edit Offer | |
|---------------------|---|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to edit the information of the Offer |
| Pre-Conditions | Offer ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Offer is modified in the system, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Offers from Admin Main Page2. All Offers are appeared in the web page3. Admin select EDIT Offers option of the specific advertiser from the web page4. Admin edit the Offers according to its need and click save button5. System updates |
| Alternative courses | Not all mandatory data fields are filled |

Detail Offers

Table 10: Use Case Detail Offers

| Detail Offer | |
|----------------|---|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to view Detail information of the Offer |
| Pre-Conditions | Offer ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |

| Detail Offer | |
|---------------------|--|
| Element | Details |
| Post Conditions | Offer is modified in the system, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Offers from Admin Main Page2. All Offer are appeared in the web page3. Admin select Details Offer option of the specific Offer from the web page4. Admin can now select View Detailed information about Offer |
| Alternative courses | Admin select Back To List Option to View All Offer |

Approve Offers

Table 11: Use Case Approve Offers

| Approve Offer | |
|---------------------|---|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to Approve Offer which will be access able to users |
| Pre-Conditions | Offer ID is not listed in the system, Offers is not listed in the approved offers, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Offer is listed in the approved offers in the system, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Offers from Admin Main Page2. All Offer are appeared in the web page3. Admin select Add Offer option of the specific Offer from the web page to Add in offer Approved4. Admin can now edit information before adding offer by Using Systematic Implementation and NLP Analysis |
| Alternative courses | Offer is already Added in the Approved List |

Delete Offers

Table 12: Use Case Delete Offers

| Delete Offer | |
|---------------------|--|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to delete the Offer |
| Pre-Conditions | Offer ID is not listed in the system, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Offer is deleted from the system, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select Offers from Admin Main Page2. All Offers are appeared in the web page3. Admin select Delete Offer option of the specific advertiser from the web page4. Admin conform the deletion of the Offer5. System updates |
| Alternative courses | Admin does not conform the deletion |

Words Analysis Cloud

Table 13: Use Case Analysis Cloud

| Word Analysis Cloud | |
|---------------------|---|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to Perform words Analysis on data Stored in the Database |
| Pre-Conditions | Data Stored in the System listed in the system, the user is logged into the system, and the system menu is displayed. |
| Post Conditions | Admin perform analysis on data from the system, the user is logged into the system, and the system menu is displayed. |
| Normal course | <ol style="list-style-type: none">1. Admin select KeyWord Cloud from Admin Main Page2. Able to see all key words |
| Alternative courses | Analysis not present |

Scrapper

Table 14: Use Case Scrapper

| Scrapper | |
|---------------------|---|
| Element | Details |
| Actor | Admin |
| Trigger | Admin wants to Scrap a web page to get data from site |
| Pre-Conditions | User is logged into the system, and the system menu is displayed. |
| Post Conditions | User is logged into the system, and the system menu is displayed and scrapped result stored in text file on you system. |
| Normal course | 1. Admin select Scraper from Admin Main Page 2. Add web page link and file name for the text file 3. Click Save |
| Alternative courses | Not all mandatory data fields are filled |

View Advertisers

Table 15: Use Case View Advertisers

| View Advertisers | |
|---------------------|--|
| Element | Details |
| Actor | User |
| Trigger | User want to View Advertisers |
| Pre-Conditions | The system menu is displayed. |
| Post Conditions | Result is displayed where user can view specific advertiser's offers and products. |
| Normal course | 1. User on the home page of site 2. On the menu bar select Advertisers and select the required sorted from of the advertisers |
| Alternative courses | Data not present in the system |

View Products

Table 16: Use Case View Products

| View Products | |
|---------------|---------|
| Element | Details |
| Actor | User |

| View Products | |
|---------------------|--|
| Element | Details |
| Trigger | User want to View Products |
| Pre-Conditions | The system menu is displayed. |
| Post Conditions | Result is displayed |
| Normal course | 3. User on the home page of site 4. On the menu bar select Products and select the required sorted from of the Products |
| Alternative courses | Data not present in the system |

View Offers

Table 17: Use Case View Offers

| View Offers | |
|---------------------|--|
| Element | Details |
| Actor | User |
| Trigger | User want to search Offers |
| Pre-Conditions | The system menu is displayed. |
| Post Conditions | Result is displayed |
| Normal course | 1. User on the home page of site 2. On the menu bar select Offers and select the required sorted from of the Offers |
| Alternative courses | Data not present in the system |

Search

Table 18: Use Case Search

| Search | |
|---------|---|
| Element | Details |
| Actor | User |
| Trigger | User want to Search Offers, Products, Advertisers by key word |

| Search | |
|---------------------|--|
| Element | Details |
| Pre-Conditions | The system menu is displayed. |
| Post Conditions | Result is displayed |
| Normal course | 1. User on the home page of site 2. On the menu bar select offer, product or advertiser and enter keyword you want to search 3. Click Search |
| Alternative courses | Data not present in the system |

Design Description

System Sequence Diagram

User System Sequence Diagram

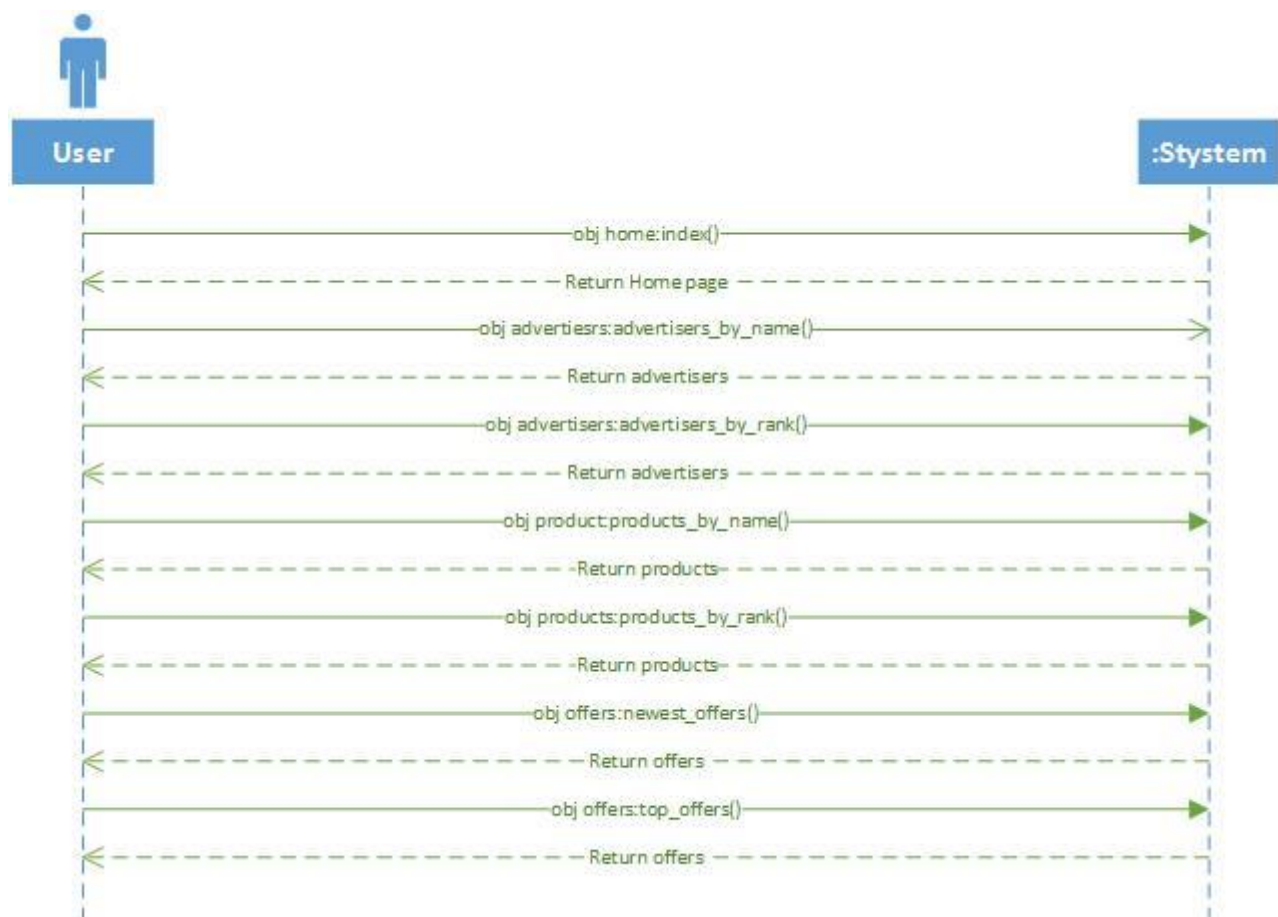


Figure 9: SSD User

Admin System Sequence Diagram



Figure 10: SSD Admin

Sequence Diagrams

Add Offers Sequence Diagram

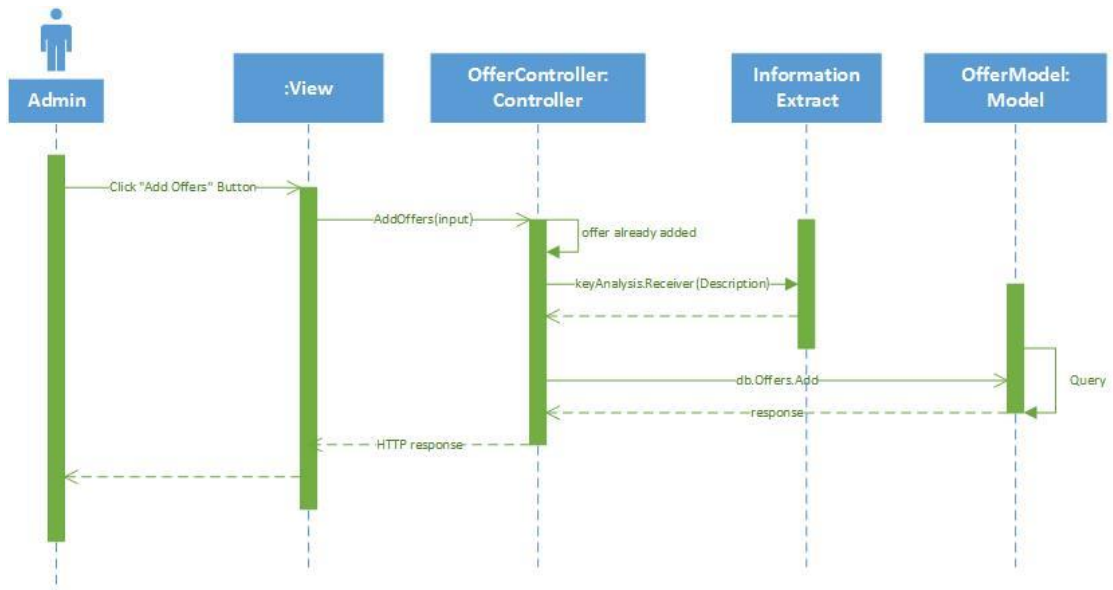


Figure 11: Add Offers Sequence Diagram

Edit Offers Sequence Diagram

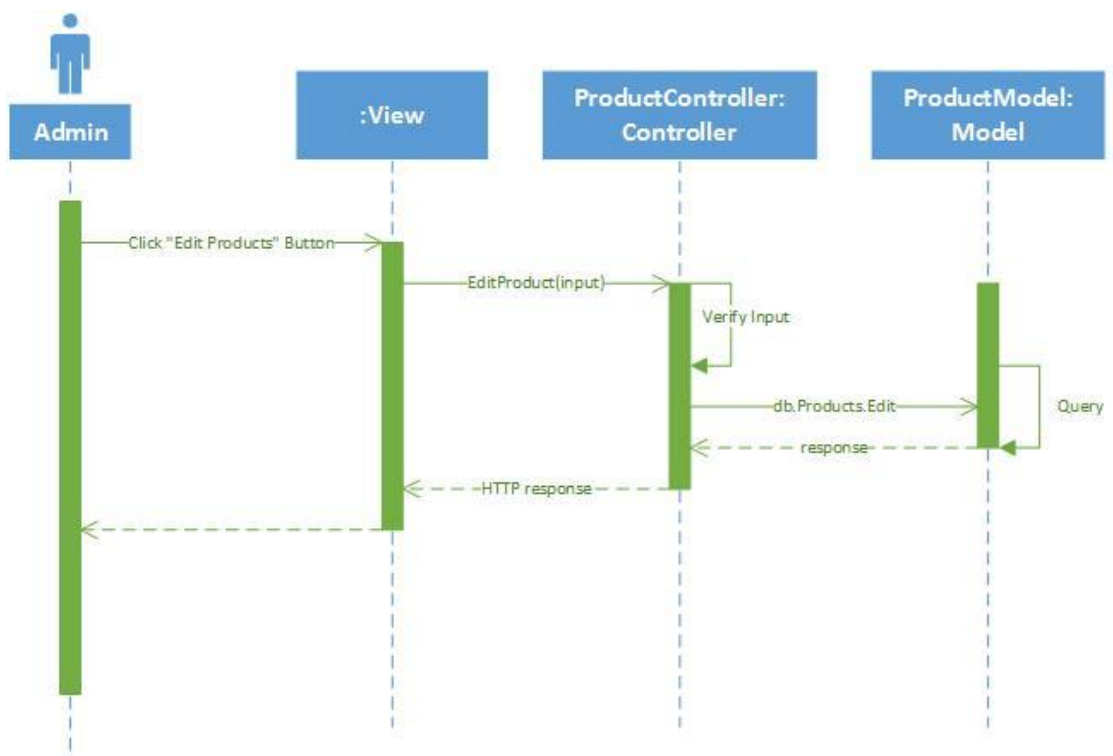


Figure 12: Edit Offers Sequence Diagram

Edit Advertisers Sequence Diagram

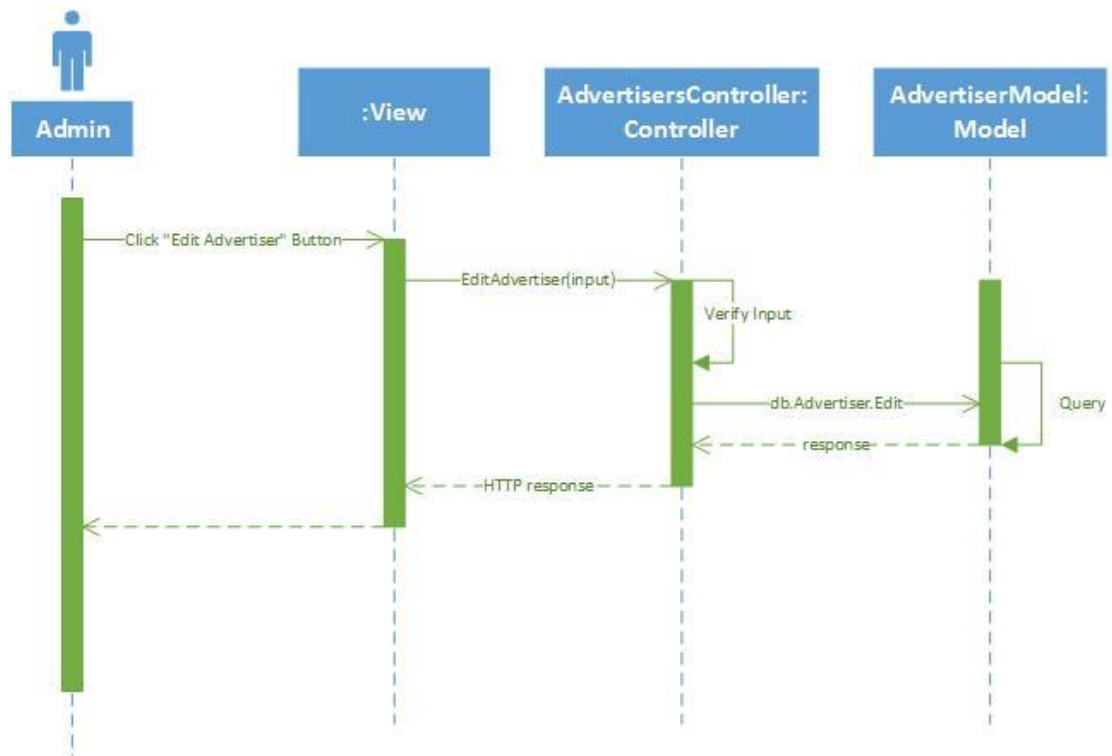


Figure 13: Edit Advertiser Sequence Diagram

Edit Offers Sequence Diagram

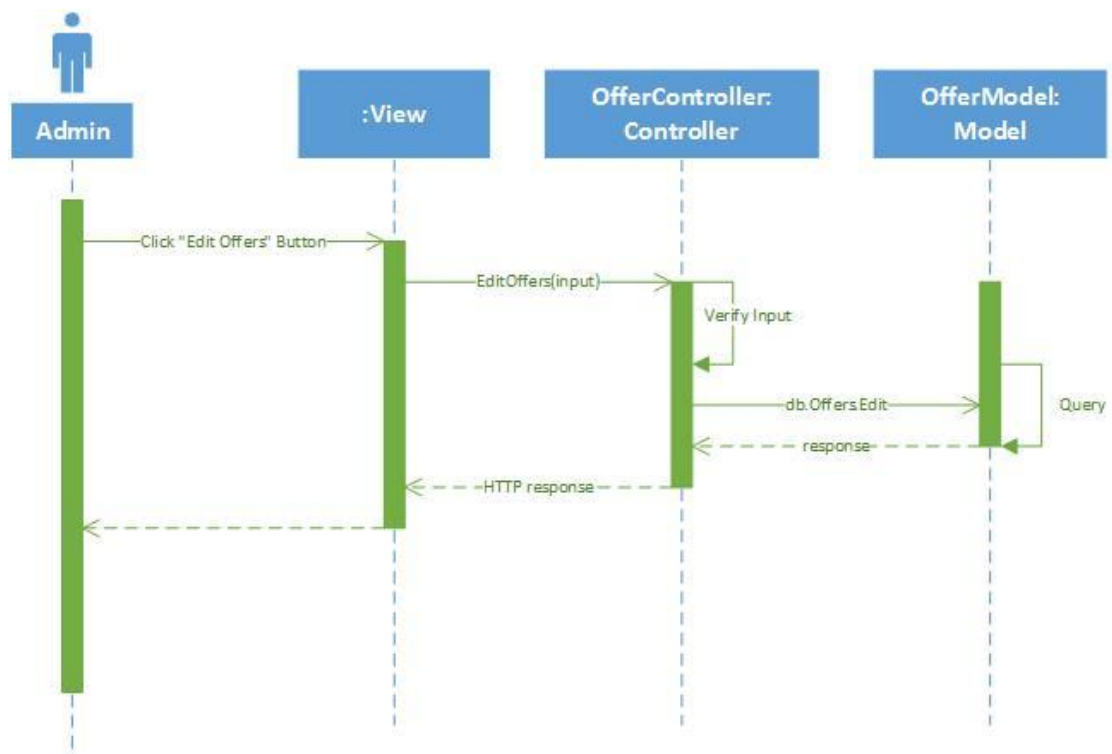


Figure 14: Edit Offers Sequence Diagram

View Offers Sequence Diagram

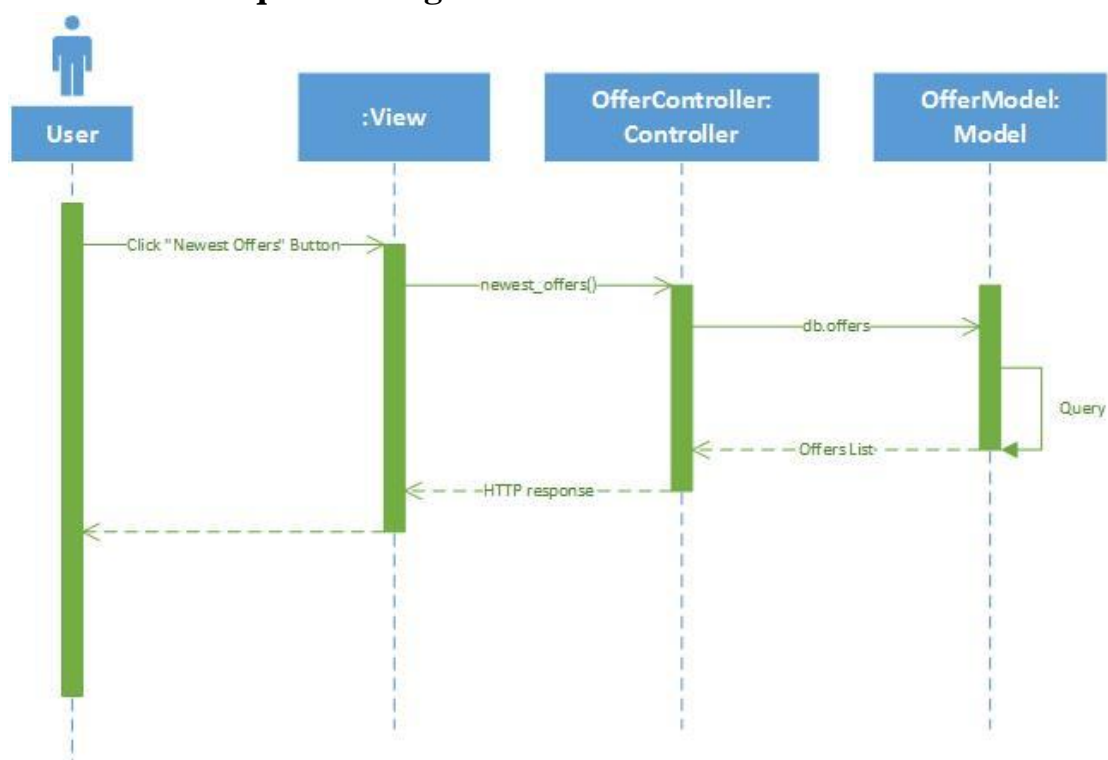


Figure 15: View Offers Sequence Diagram

View Products Sequence Diagram

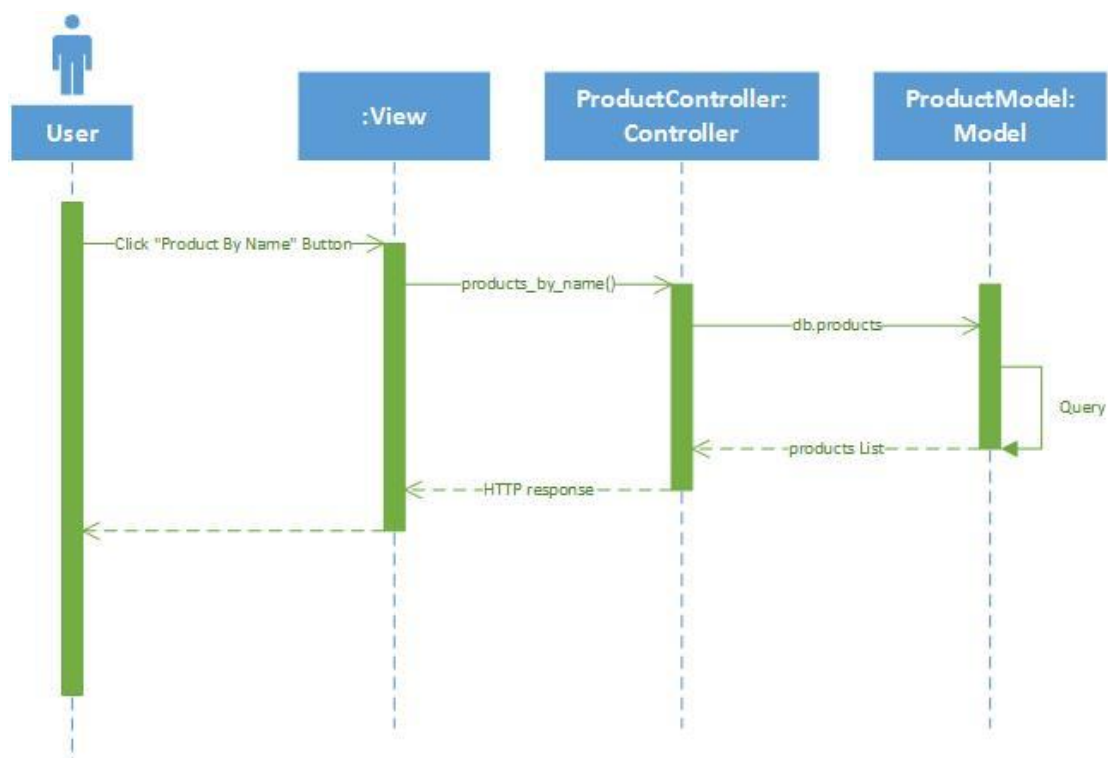


Figure 16: View Products Sequence Diagram

View Advertisers Sequence Diagram

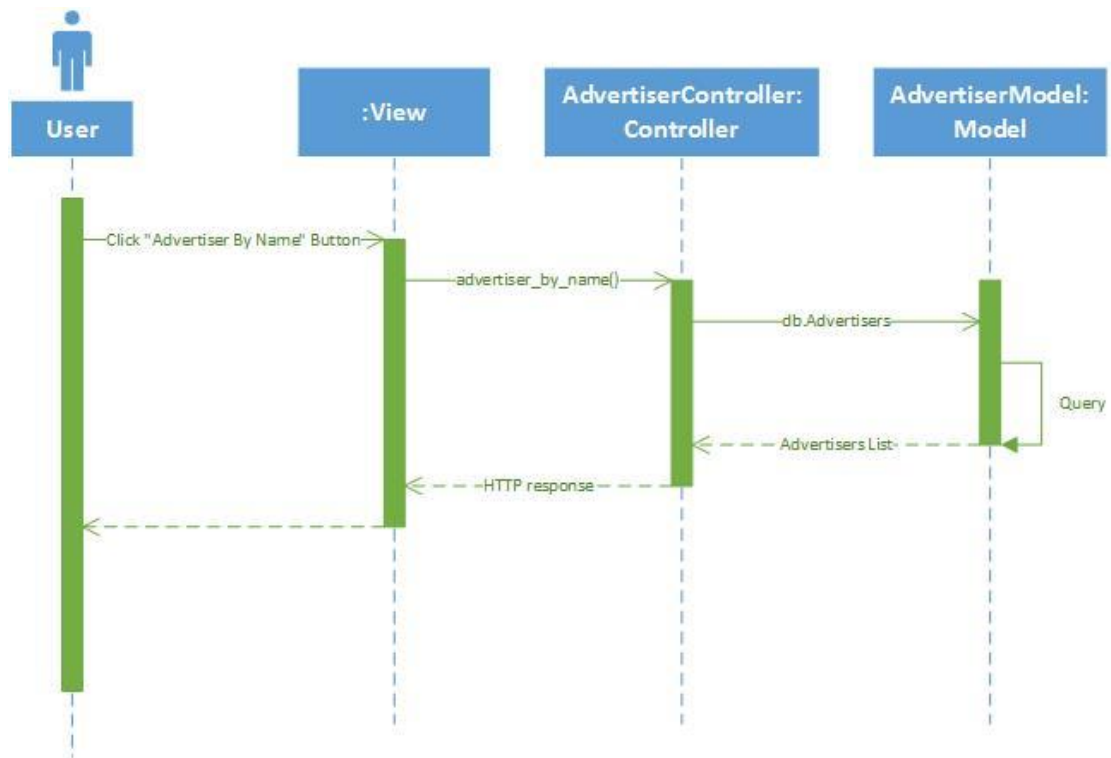


Figure 17: View Advertisers Sequence Diagram

Entity Relationship Diagram

Complete ERD

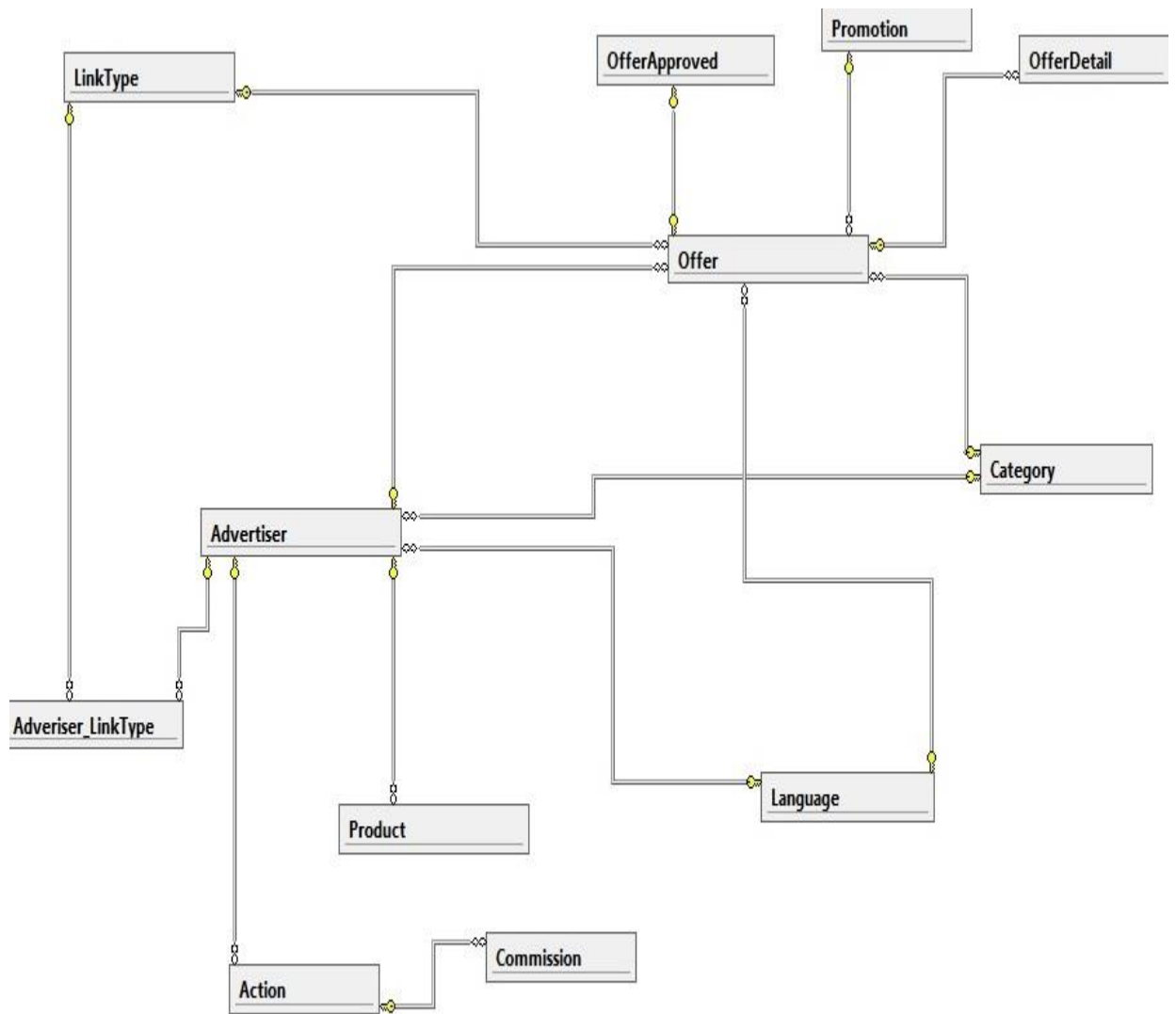


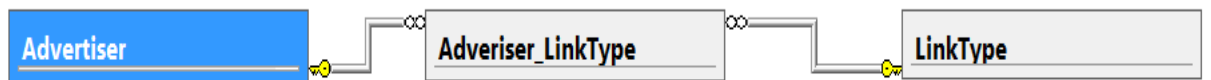
Figure 18: ERD

Unit level Relationships

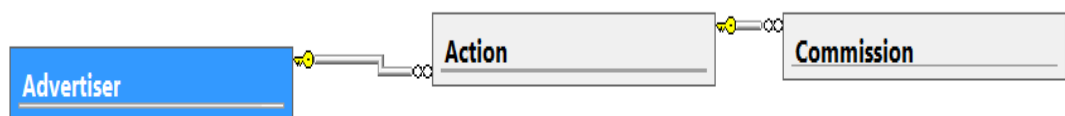
Advertiser language



Advertiser Link Type



Advertiser Action



Advertiser Category



Offer Advertiser



Offer Approved



Offer Category



Offer Detail



Offer Language



Offer Link



Offer Promotion



Product Advertiser



Data Tables

Action

Table 19: Data Table Action

| Column name | Data Type | Allow null |
|------------------|--------------|------------|
| | | |
| ActionID (PK) | int | No |
| Name | Nchar(100) | yes |
| AdvertiserID(FK) | Int | Yes |
| Type | Nchar(20) | Yes |
| Defaults | Nchar(15) | Yes |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Advertiser_LinkType

Table 20: Data Table Advertiser_LinkType

| Column name | Data Type | Allow null |
|------------------|--------------|------------|
| | | |
| AdvertiserID(FK) | int | yes |
| LinkTypeID(FK) | int | yes |
| ID(PK) | Int | no |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Advertiser

Table 21: Data Table Advertiser

| Column name | Data Type | Allow null |
|----------------------|--------------|------------|
| | | |
| AdvertiserID(PK) | int | No |
| AccountStatus | Nchar(15) | yes |
| SevenDayEPC | Floar | Yes |
| ThreeMonthEPC | float | Yes |
| LanguageID(FK) | int | Yes |
| Name | Nchar(80) | Yes |
| URL | Nchar(150) | Yes |
| RelationShipStatus | Nchar(15) | Yes |
| MobileTracking | Nchar(10) | Yes |
| NetworkRank | Nchar(10) | Yes |
| ParentCategoryID(FK) | Int | Yes |
| ChildCategoryID(FK) | Int | Yes |
| Performancelncentive | Nchar(10) | Yes |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Category

Table 22: Data Table Category

| Column name | Data Type | Allow null |
|----------------|--------------|------------|
| | | |
| Category | Nchar(50) | yes |
| CategoryID(PK) | Int | no |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Commission

Table 23: Data Table Commission

| Column name | Data Type | Allow null |
|----------------|--------------|------------|
| | | |
| ActionID(FK) | int | yes |
| ItemListID(PK) | int | No |
| ItemListName | Nchar(120) | Yes |
| PricePercent | Nchar(15) | Yes |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Language

Table 24: Data Table Language

| Column name | Data Type | Allow null |
|----------------|--------------|------------|
| | | |
| LanguageName | Nchar(3) | yes |
| LanguageID(PK) | int | No |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

LinkType

Table 25: Data Table LinkType

| Column name | Data Type | Allow null |
|----------------|--------------|------------|
| | | |
| LinkType | Nchar(50) | yes |
| LinkTypeID(PK) | int | no |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Offer

Table 26: Offer

| Column name | Data Type | Allow null |
|----------------------|---------------|------------|
| | | |
| AdvertiserID(FK) | int | Yes |
| CategoryID(FK) | int | Yes |
| ClickCommission | float | Yes |
| LanguageID(FK) | Int | Yes |
| LeadCommission | Nchar(40) | Yes |
| Description | Nvarchar(Max) | Yes |
| Destination | Nvarchar(Max) | Yes |
| LinkID | Int | Yes |
| LinkTypeID(FK) | int | Yes |
| PerformanceIncentive | Nchar(10) | Yes |
| PromotionEndDate | Datetime2(7) | Yes |
| PromotionStartDate | Datetime2(7) | Yes |
| promotionID(FK) | int | Yes |
| CouponCode | Nchar(50) | Yes |
| RelationStatus | Nchar(10) | Yes |
| SalesCommission | Nchar(60) | Yes |
| SevenDayEPC | Nchar(10) | Yes |
| ThreeMonthEPC | Nchar(10) | Yes |
| ClickURL | Nvarchar(MAX) | Yes |
| LinkName | Nvarchar(MAX) | Yes |

| | | |
|-------------|--------------|-----|
| OfferID(PK) | int | No |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

OfferDetail

Table 27: Data Table OfferDetail

| Column name | Data Type | Allow null |
|--------------------|---------------|------------|
| | | |
| HTMLCode | Nvarchar(max) | yes |
| JavaScriptCode | Nvarchar(max) | yes |
| Heigh | int | yes |
| Width | Int | yes |
| OfferID(fk) | Int | No |
| OfferDetailKey(PK) | Int | no |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Promotion

Table 28: Data Table Promotion

| Column name | Data Type | Allow null |
|-----------------|--------------|------------|
| | | |
| PromotionType | Nchar(25) | Yes |
| PromotionID(PK) | int | no |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Product

Table 29: Data Table Product

| Column name | Data Type | Allow null |
|--------------------|---------------|------------|
| | | |
| AdID | int | Yes |
| AdvertiserID(FK) | int | Yes |
| AdvertisorCategory | Nchar(275) | Yes |
| BuyURL | Nvarchar(max) | Yes |
| CatalogID | Nchar(12) | Yes |
| Currency | Nchar(5) | Yes |
| Description | Nvarchar(Max) | Yes |
| ImageUrl | Nvarchar(Max) | Yes |
| InStock | Nchar(10) | Yes |
| ManufacturerSKU | Nchar(110) | Yes |
| Name | Nchar(400) | Yes |
| Price | Float | Yes |

| | | |
|------------------|--------------|-----|
| RetailPrice | Float | Yes |
| SalePrice | float | Yes |
| SKU | Nchar(150) | Yes |
| UPC | Nchar(50) | Yes |
| ISBN | Nchar(40) | Yes |
| ProductKey(PK | Int | NO |
| ManufacturerName | Nchar(150) | Yes |
| CreateDate | Datetime2(7) | Yes |
| ModifyDate | Datetime2(7) | Yes |

Class Diagram

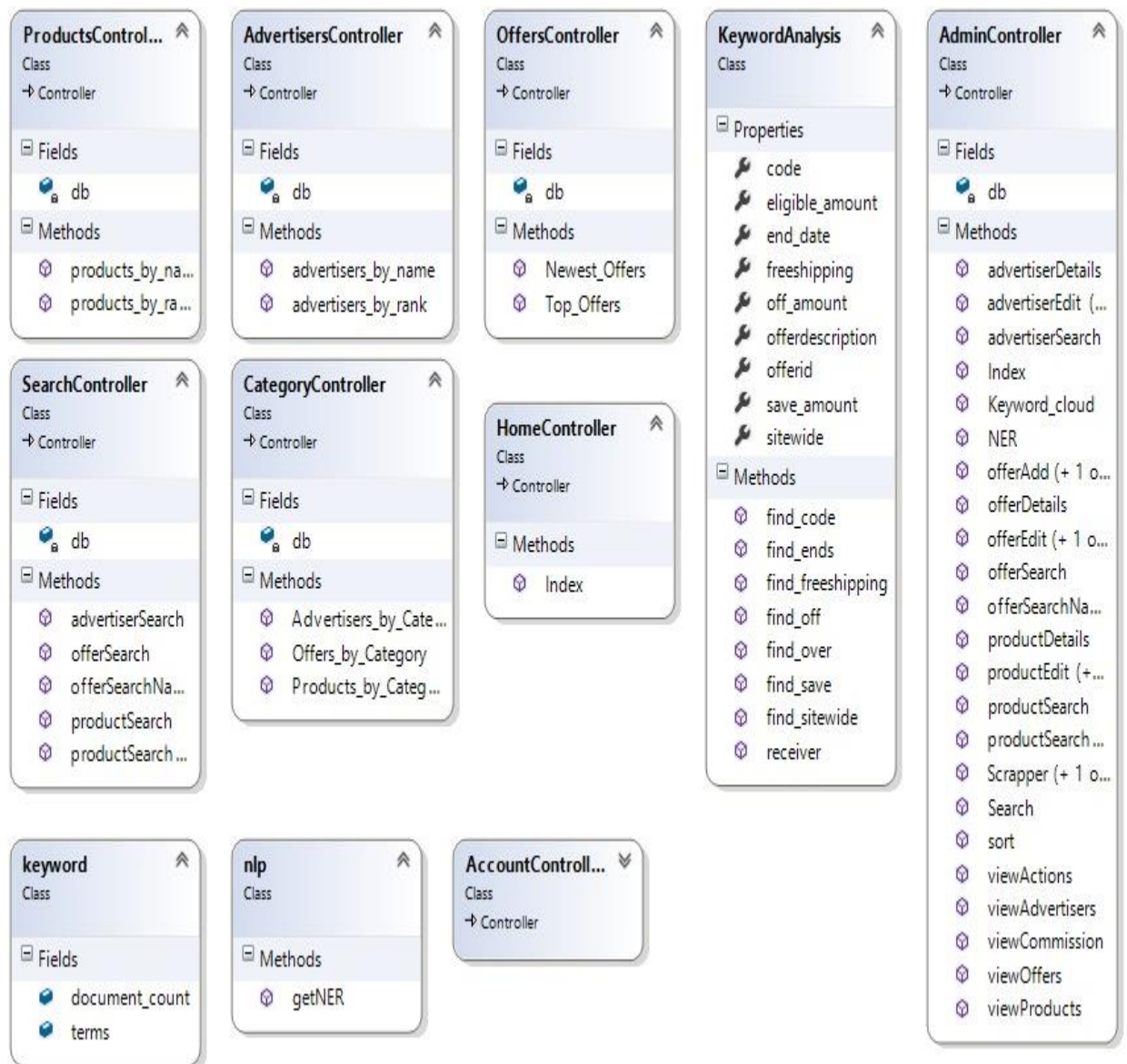


Figure 19: Class Diagram

Activity Diagram

Admin Activity Diagram

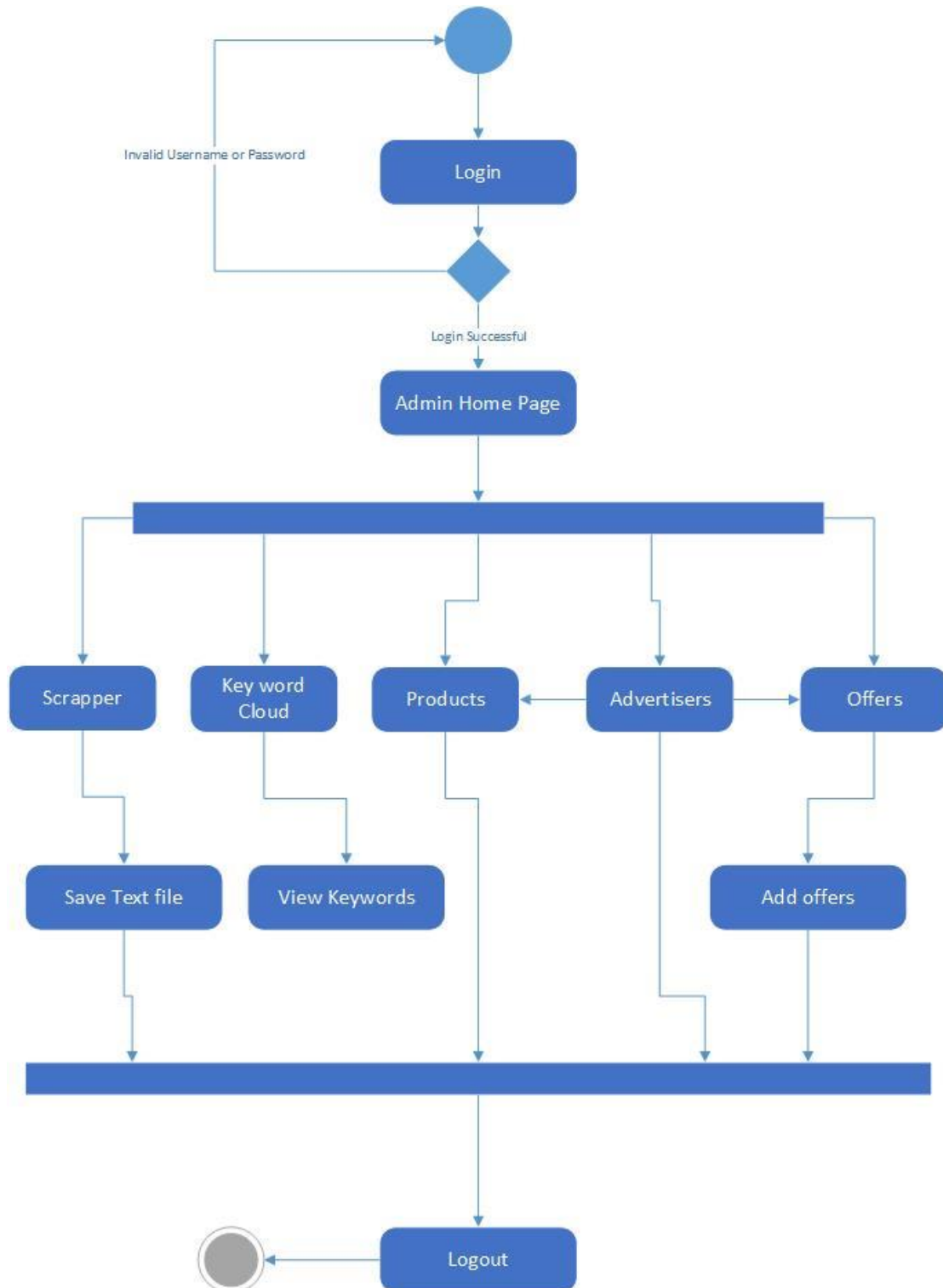


Figure 20: Activity Diagram

Approve Offer

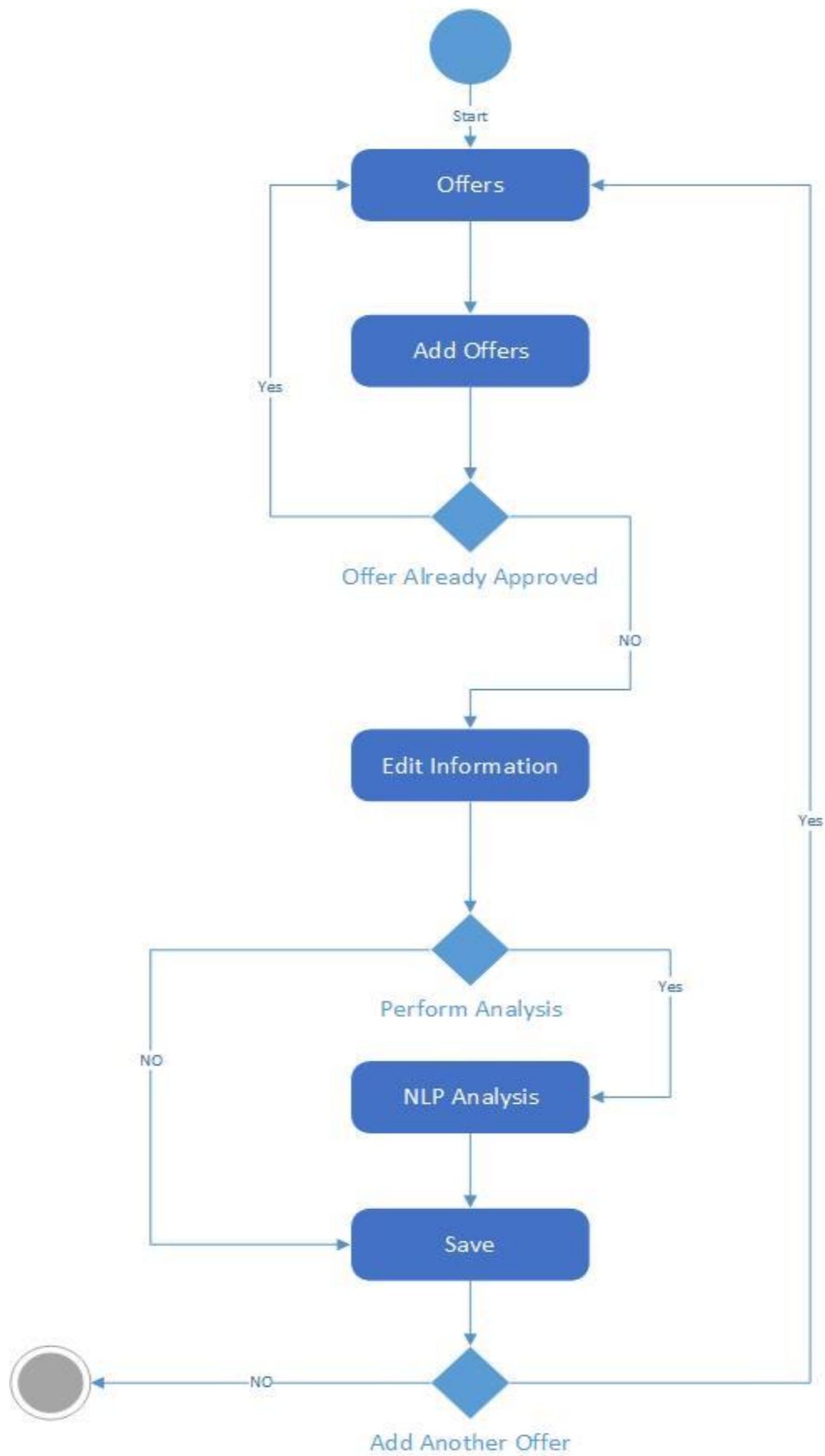


Figure 21: Approve Offers

Advertiser Detail

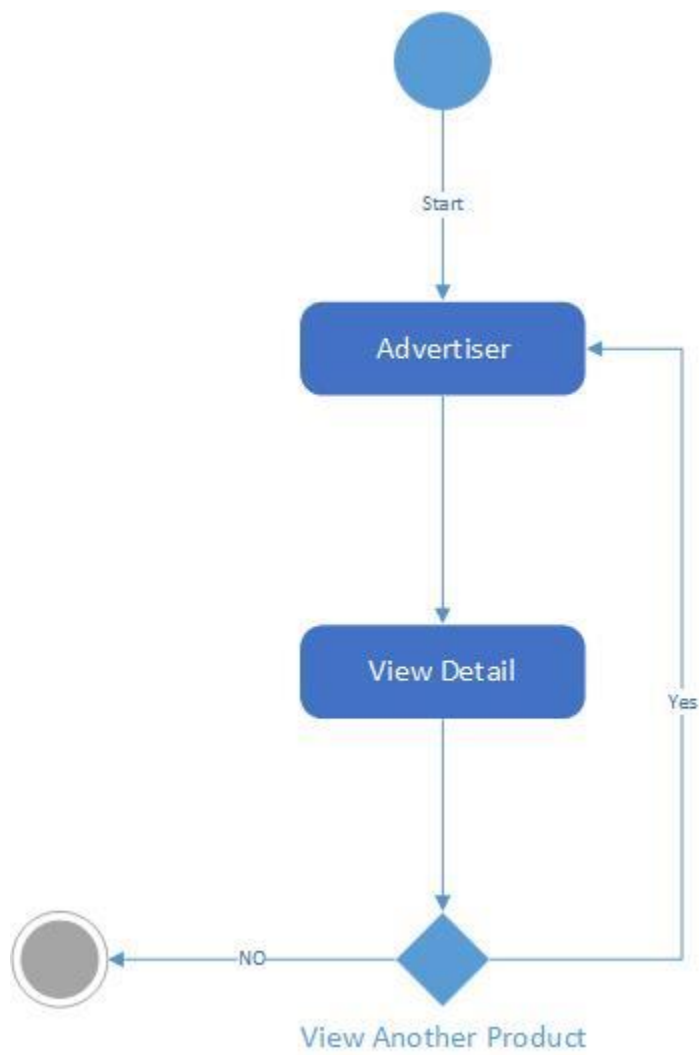


Figure 22: Advertiser Detail

Offer Detail

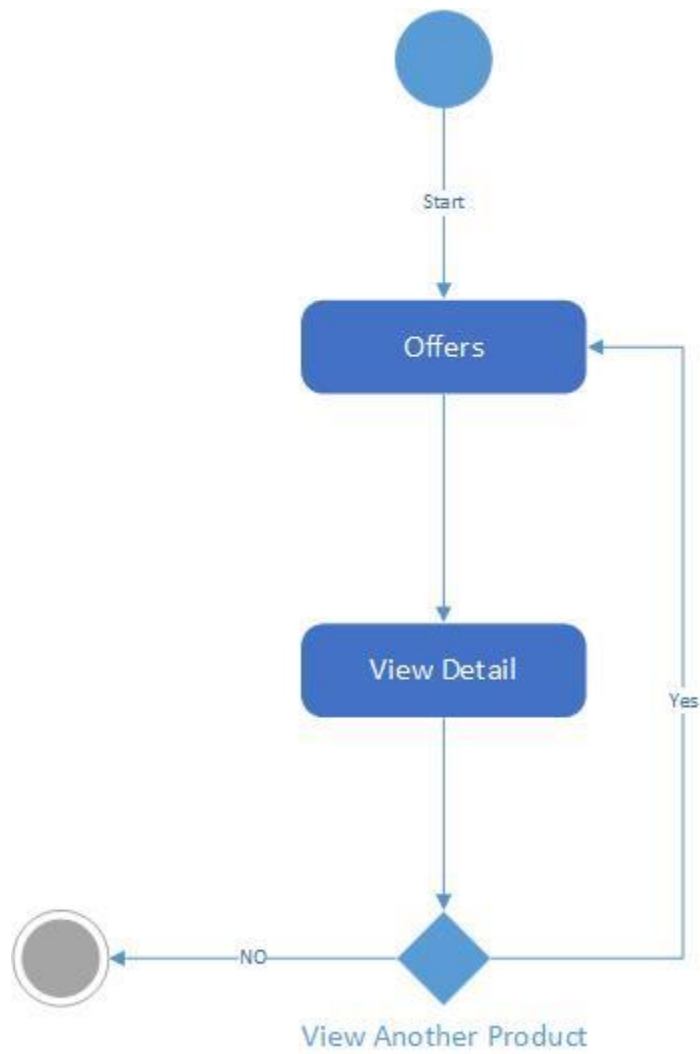


Figure 23: Offer Detail

Product Detail

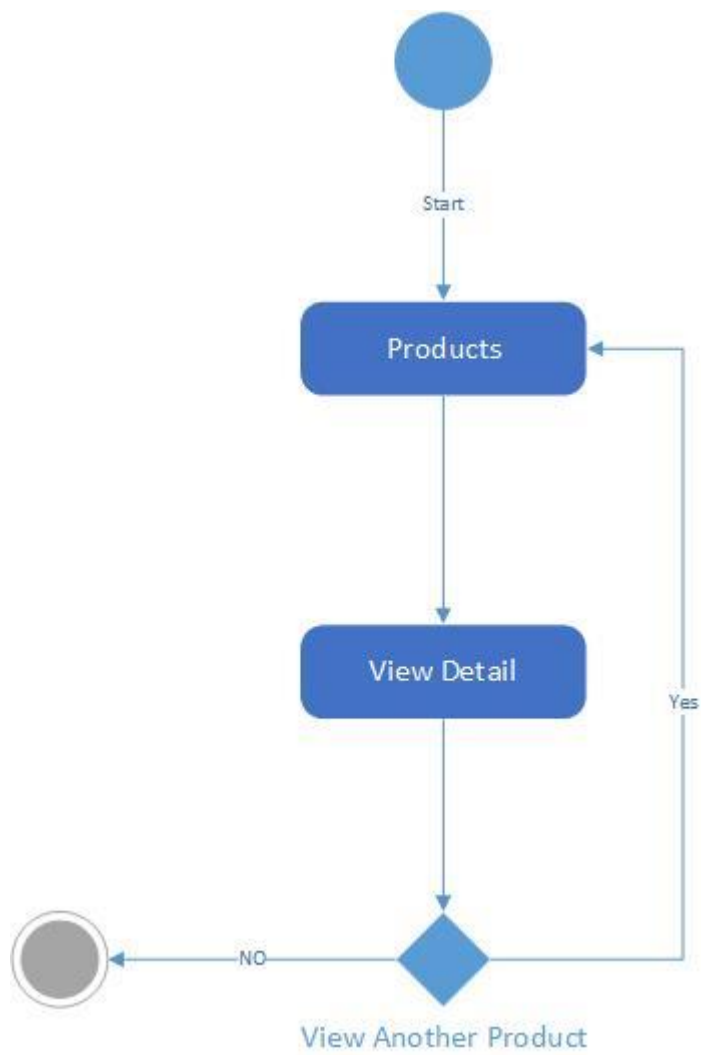


Figure 24: Product Detail

Advertiser Edit

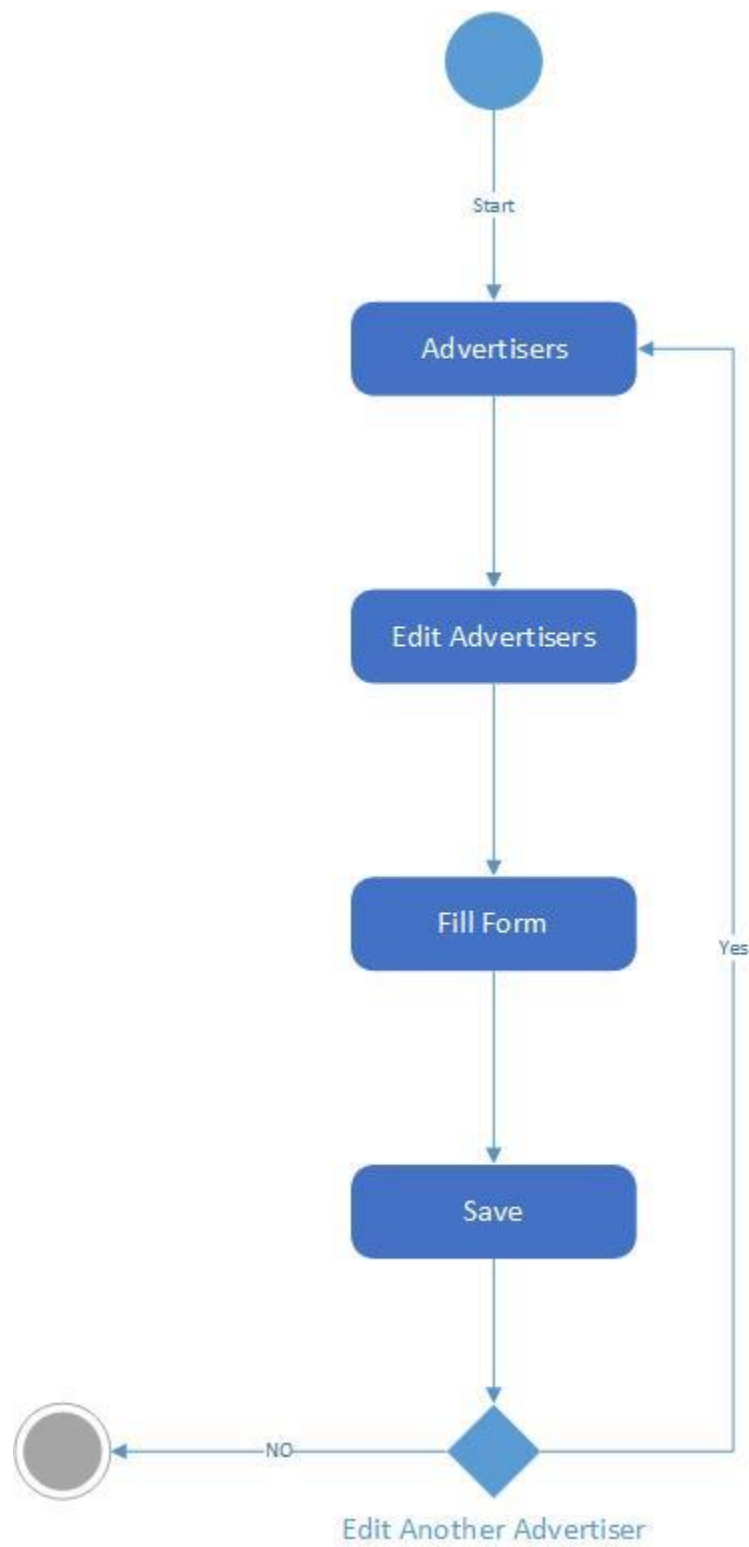


Figure 25: Advertiser Edit

Edit Offer

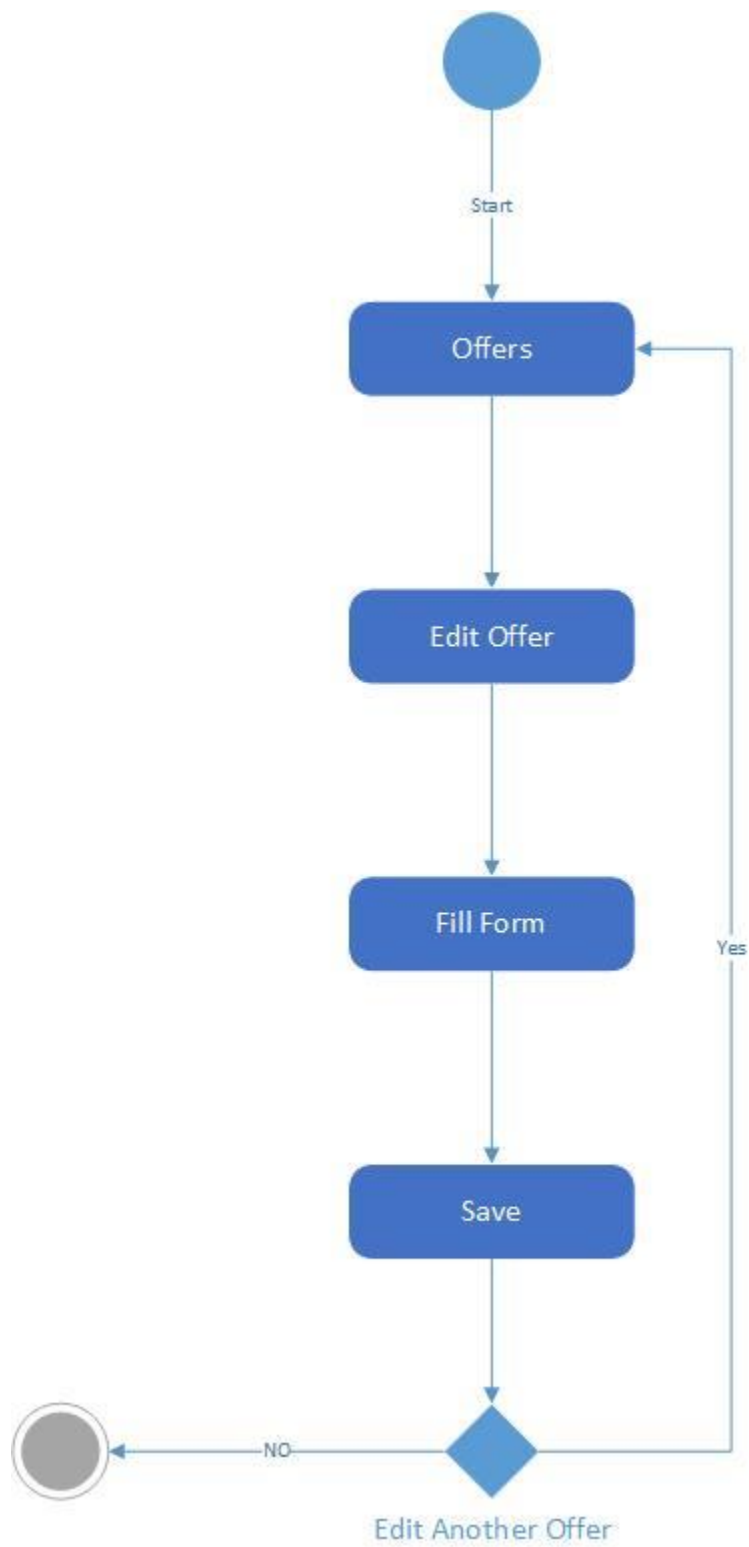


Figure 26: Offer Edit

Edit Product

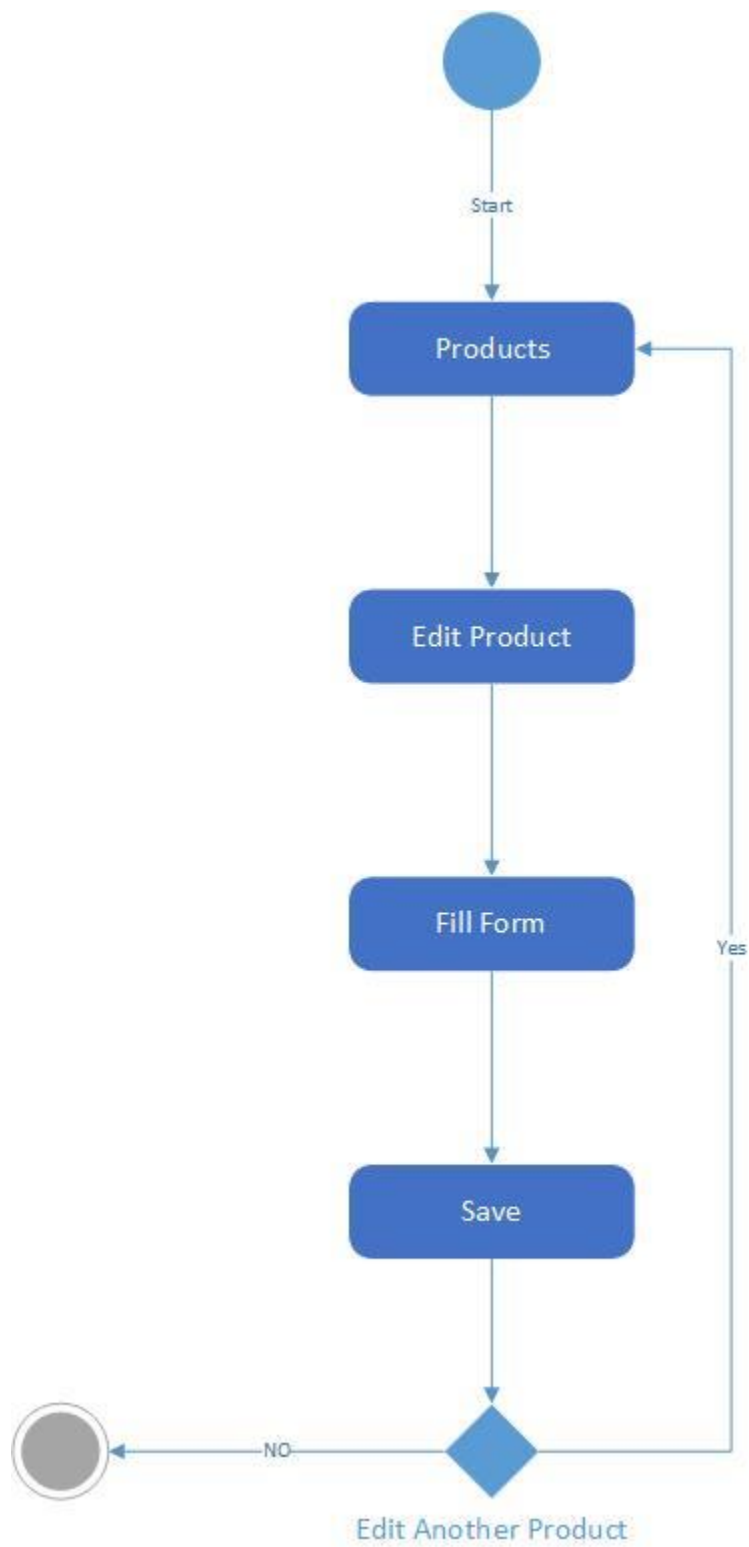


Figure 27: Edit Product

IMPLEMENTATION

Development Stages

Research

Research was an umbrella task during our whole project but in start we did some research as a separate task. During our initial research we started looking into SEO, we looked into a tool named as IBP (Internet Business Promoter). It is a tool that helps you improving your search ranking and by using it we got a good idea that how SEO works. After that we looked into google algorithms and studied its Panda, Penguin, Humming Bird and Brain rank algorithms. At the end of it we started studying schem.org and nlp.

Planning

We applied evolutionary software development approach to plan our project. We decided that first we should focus on information extraction and later implementation would be done later.

Design

In design phase we made our database structure. Decided the actors and their roles. We also looked upon the hierarchy or flow of controls. Because our development approach was evolutionary so designing phase saw a lot of changes.

Development:

We developed our information extraction module as a first priority. First we try to do it only with nlp and then later on inducted keyword analysis to make it more reliable. Later on we developed our site and integrated these systems into it.

The tools and frameworks we used in development of our project are as follows

- Visual studio 2015
- .net framework 4.6
- Entity Framework 6.0
- MVC 5.0
- Asp.net Razor
- Sql Server 2012

Testing

After the development we tested the site by test cases given in this document moreover we also conducted a lot of other white box tests that we didn't mention here.

User Interface

Following is the user interface of our project

Main Screen

The main screen is the first screen that a user see on the site. From here user can navigate to any other page. From side bar he can select category and see advertisers, products and offers on that particular category. From top menu he can select advertisers, products and offers and view them according to its choice. From the search bar it can search products, advertisers or offers that he is trying to find. At the

bottom of the screen there is a login option for admin.

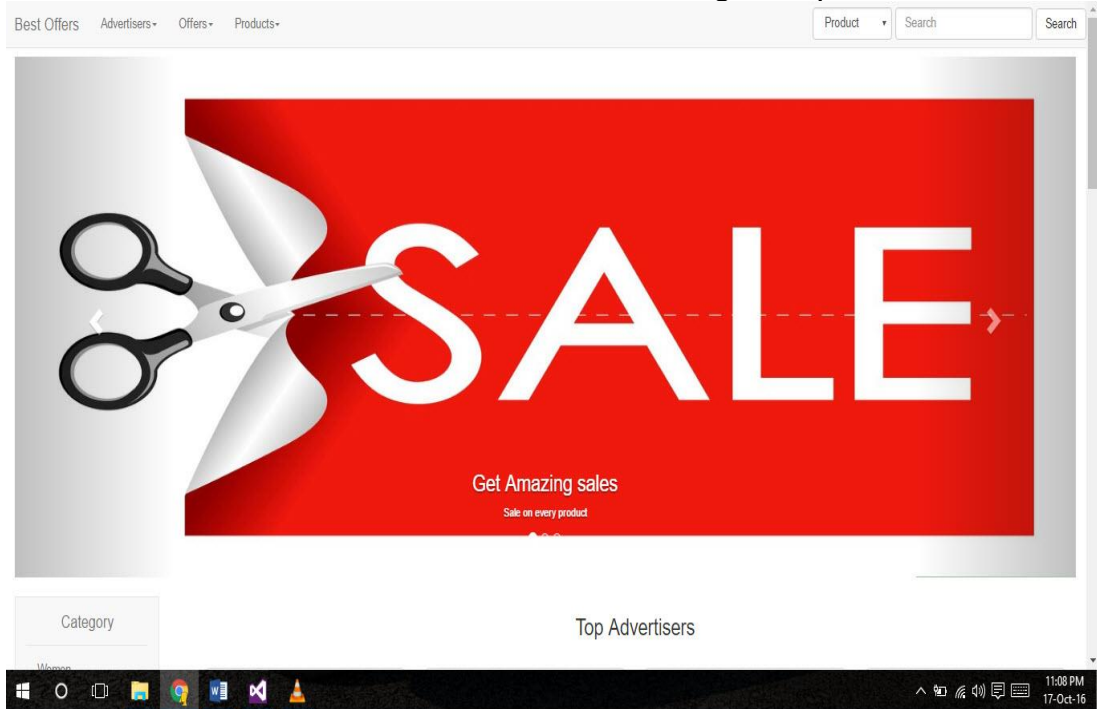


Figure 28: Home Page

Advertiser screen

Advertiser screen contains different advertisers that could be sorted alphabetically or by their rank by selecting an advertiser user can view its products and offers. From this screen user can also go to next page to view more advertisers

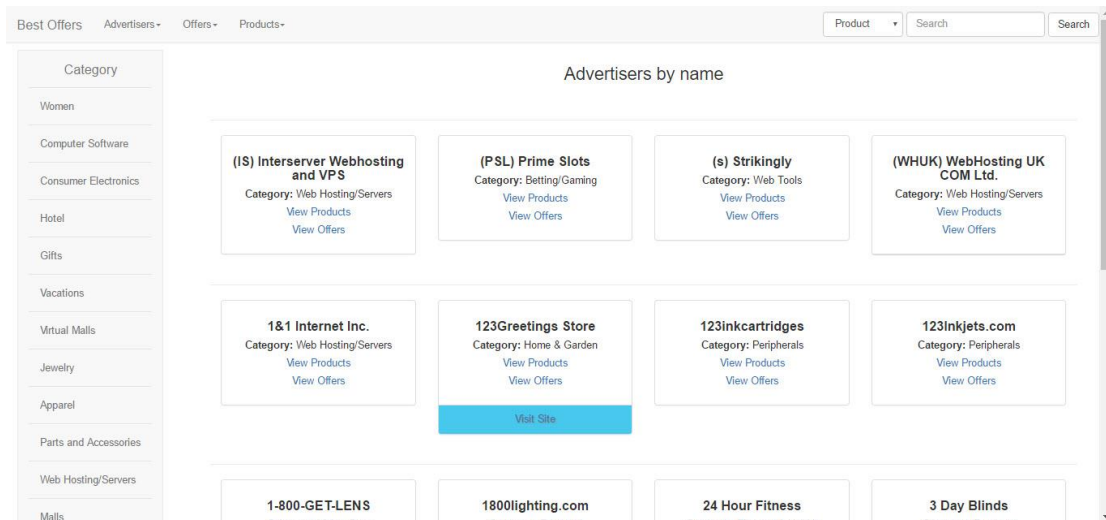


Figure 29: Advertiser Screen

Product Screen

From this screen user can see different products their price, manufacturer and other details as well

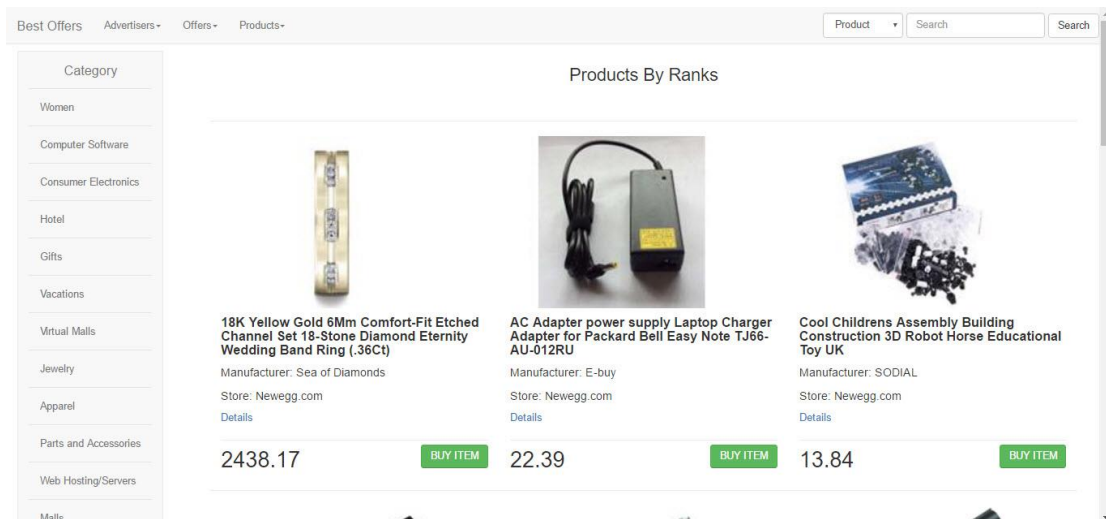


Figure 30: Product Screen

Offer Screen

From offer page user can view different offers and details corresponding to it. E.g start date, end date discount etc

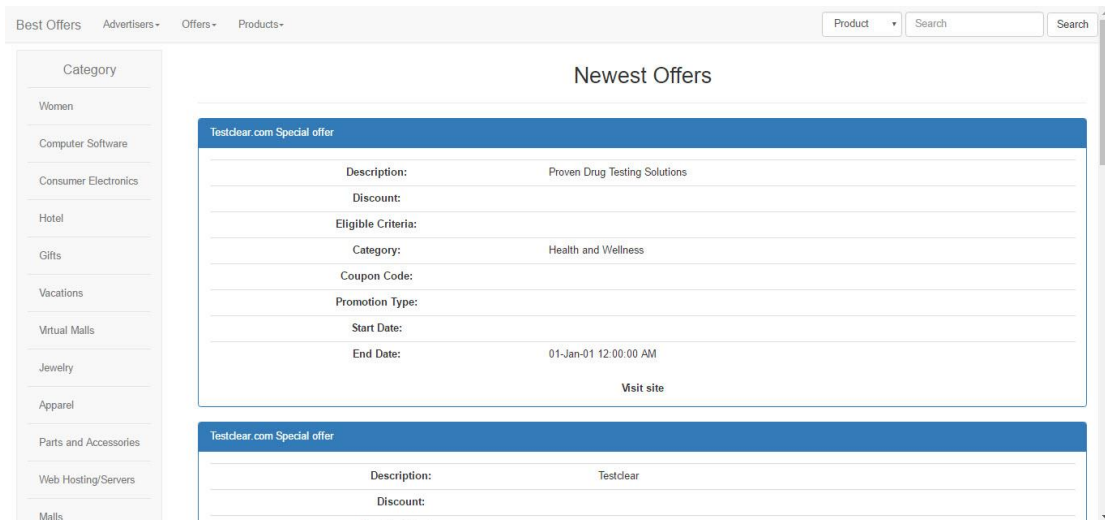


Figure 31: Offer Screen

Login Screen

Login Screen is there so that admin could login and access admin side

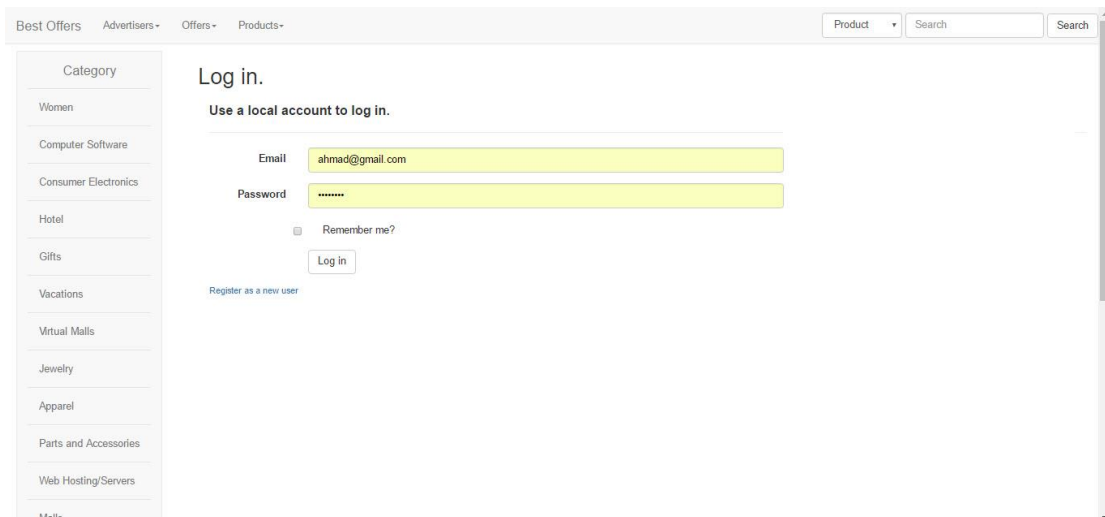
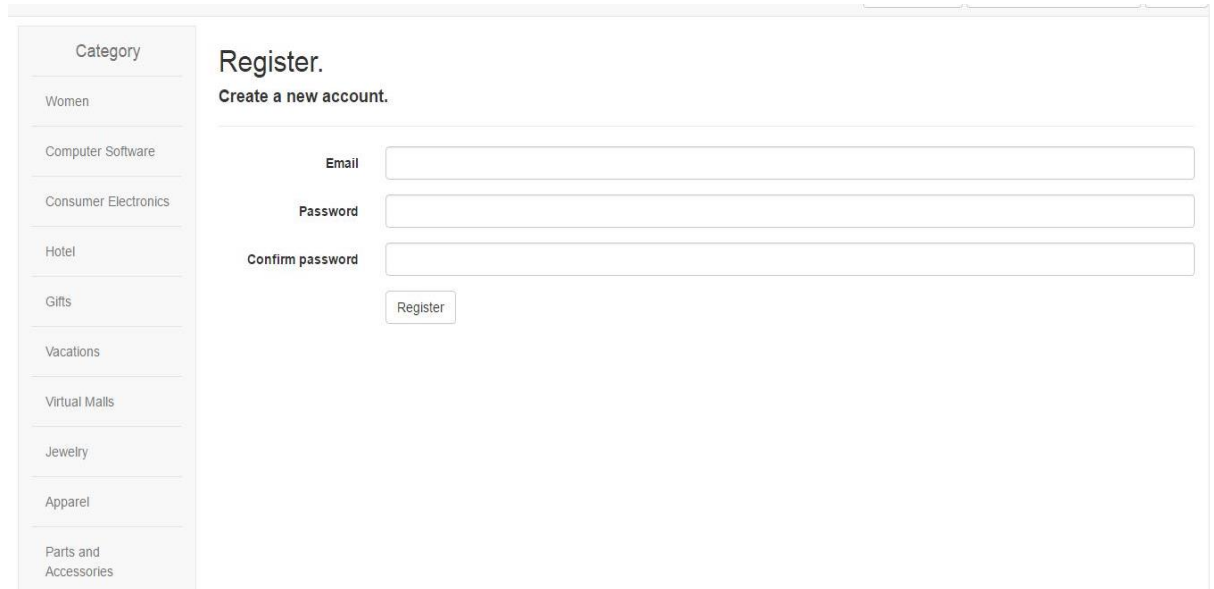


Figure 32: Login Screen

Register new admin

Register new admin screen is there so that new admin could be registered. Note only previously registered admin can register a new admin

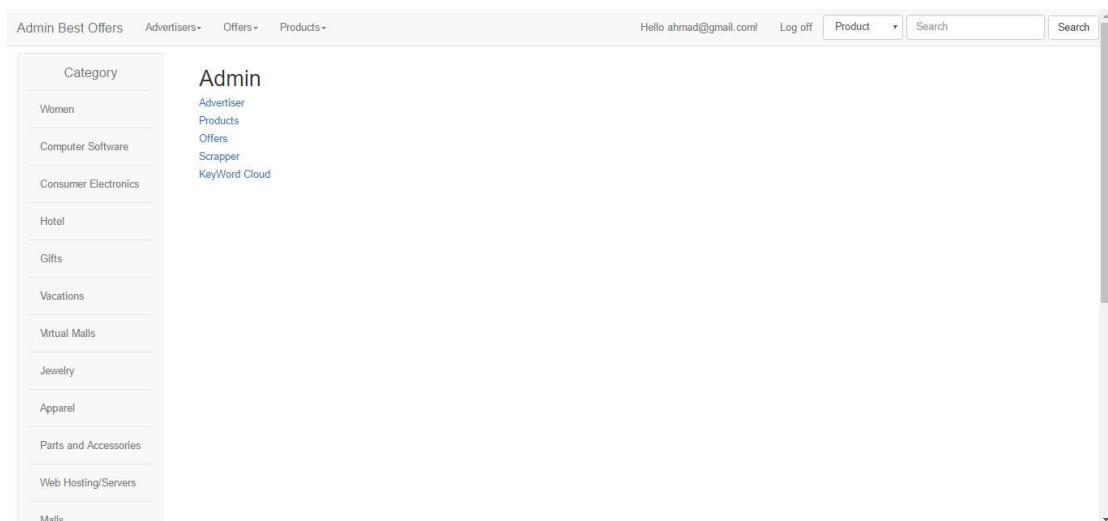


The Register screen features a sidebar on the left with a 'Category' header and a list of categories: Women, Computer Software, Consumer Electronics, Hotel, Gifts, Vacations, Virtual Malls, Jewelry, Apparel, and Parts and Accessories. The main content area is titled 'Register.' and includes the subtext 'Create a new account.' Below this, there are three input fields labeled 'Email', 'Password', and 'Confirm password'. A 'Register' button is positioned below the 'Confirm password' field.

Figure 33: Register Screen

Main Admin Screen

Admin Screen acts as a main entry point for admins and it allow admin to access different admin side modules



The Admin Home Page has a top navigation bar with links for 'Admin Best Offers', 'Advertisers', 'Offers', and 'Products'. On the right side of the top bar, it displays 'Hello ahmad@gmail.com', a 'Log off' link, a 'Product' dropdown menu, and a search bar with a 'Search' button. The main content area is titled 'Admin' and lists several modules: 'Advertiser', 'Products', 'Offers', 'Scraper', and 'KeyWord Cloud'. A sidebar on the left contains a 'Category' header and a list of categories: Women, Computer Software, Consumer Electronics, Hotel, Gifts, Vacations, Virtual Malls, Jewelry, Apparel, Parts and Accessories, Web Hosting/Servers, and Malls.

Figure 34: Admin Home Page

Admin Advertiser screen

This screen allows admin to perform update operations upon admin.

| Name | SevenDayEpc | ThreeMonthEpc | Language | AccountStatus | RelationshipStatus | MobileTracking | NetworkRank | ParentCategoryID | ChildCategoryID |
|--|-------------|---------------|----------|---------------|--------------------|----------------|-------------|------------------------|--------------------|
| Edit Details (IS) Interserver Webhosting and VPS | 76 | 4.26 | en | Active | joined | false | 3 | Online Services | Web Hosting/Server |
| Edit Details (PSL) Prime Slots | 4.09 | 1.79 | en | Active | joined | false | 1 | Recreation & Leisure | Betting/Gaming |
| Edit Details (s) Strikingly | 21.03 | 47.15 | en | Active | joined | false | 4 | Online Services | Web Tools |
| Edit Details (WHUK) WebHosting UK COM Ltd. | 0.54 | 5.96 | en | Active | joined | false | 0 | Online Services | Web Hosting/Server |
| Edit Details 181 Internet Inc. | 53.07 | 42.7 | en | Active | joined | false | 5 | Online Services | Web Hosting/Server |
| Edit Details 123Greetings Store | 9.42 | 2.24 | en | Active | joined | false | 2 | | Home & Garden |
| Edit Details 123inkcartridges | 33.73 | 34.82 | en | Active | joined | true | 3 | Computer & Electronics | Peripherals |
| Edit Details 173inklate.com | 7.61 | 16.74 | en | Active | joined | true | 4 | Computer & Electronics | Peripherals |

Figure 35: Advertiser Detail Screen

Admin Advertiser Edit Screen

This screen allows admin to edit advertiser details

Edit Advertiser

| | |
|--------------------|-------------------------------------|
| Name | (IS) Interserver Webhosting and VPS |
| AccountStatus | Active |
| SevenDayEpc | 76 |
| ThreeMonthEpc | 4.26 |
| LanguageID | 9 |
| Url | http://www.interserver.net |
| RelationshipStatus | joined |
| MobileTracking | false |
| NetworkRank | 3 |
| ParentCategoryID | 64 |
| ChildCategoryID | 108 |

Figure 36: Advertiser Edit Screen

Admin Advertiser Detail

This screen allows admin to see details of advertisers

Systematic Implementation of Schema.Org

The screenshot shows the 'AdvertiserDetails' screen. On the left is a category sidebar with options like Women, Computer Software, Consumer Electronics, Hotel, Gifts, Vacations, Virtual Malls, Jewelry, Apparel, Parts and Accessories, Web Hosting/Servers, and Malls. The main content area displays details for an advertiser named 'Interserver Webhosting and VPS'. The details include AccountStatus (Active), SevenDayEpc (76), ThreeMonthEpc (4.25), LanguageID (9), Name, Url (http://www.interserver.net), RelationshipStatus (joined), MobileTracking (false), NetworkRank (3), ParentCategoryID (64), ChildCategoryID (108), PerformanceIncentive (false), CreateDate (28-Aug-16 12:00:00 AM), and ModifyDate (28-Aug-16 12:00:00 AM). At the bottom, there are links for 'Actions Of Advertiser | Edit | Back to List'.

Admin Best Offers Advertisers- Offers- Products- Hello ahmad@gmail.com! Log off Product Search Search

advertiserDetails

Advertiser

[View](#) [View Products](#) | [View Offers](#)

AccountStatus: Active
SevenDayEpc: 76
ThreeMonthEpc: 4.25
LanguageID: 9
Name: (IS) Interserver Webhosting and VPS
Url: http://www.interserver.net
RelationshipStatus: joined
MobileTracking: false
NetworkRank: 3
ParentCategoryID: 64
ChildCategoryID: 108
PerformanceIncentive: false
CreateDate: 28-Aug-16 12:00:00 AM
ModifyDate: 28-Aug-16 12:00:00 AM

[Actions Of Advertiser](#) | [Edit](#) | [Back to List](#)

Figure 37: Advertiser Details Screen

Admin Product screen

This page allows admin to see products and select them to perform update operations

The screenshot shows the 'ProductsDetail' screen. On the left is the same category sidebar as in Figure 37. The main content area displays a table of products. The table has columns: Name, Price, RetailPrice, SalePrice, AdID, Advertiser, AdvertiserCategory, CatalogID, Currency, In Stock, ManufacturerSKU, SKU, UPC, and ISB. There are six rows of product data, each with an 'Edit Details' link.

Admin Best Offers Advertisers- Offers- Products- Hello ahmad@gmail.com! Log off Product Search Search

ProductsDetail

| | Name | Price | RetailPrice | SalePrice | AdID | Advertiser | AdvertiserCategory | CatalogID | Currency | In Stock | ManufacturerSKU | SKU | UPC | ISB |
|------------------------------|---|-------|-------------|-----------|----------|-------------------------------|------------------------------|-----------|----------|----------|-----------------|--------------|-----|-----|
| Edit Details | " Dreamgirl Sheer Seamed Stockings" | 5.82 | | | 11350107 | The Tight Spot | Stockings > Seamed Stockings | qjo 3059 | EUR | true | | DG022-C | | |
| Edit Details | " Pamela Mann Net Bow Suspender Tights" | 12.87 | | | 11350107 | The Tight Spot | Tights > Patterned Tights | qjo 3059 | EUR | true | | TOP2037-C | | |
| Edit Details | " Pamela Mann Net Bow Suspender Tights" | 14.02 | | | 11350108 | The Tight Spot | Tights > Patterned Tights | qjo 3101 | USD | true | | TOP2037-C | | |
| Edit Details | ""Little Notes"" Journal Blush Grey Full Grain Leather" | 67 | | | 11358494 | GiGi New York / Graphic Image | handbags & accessories | qjo 2334 | usd | true | | MTJ-NOTE-BLS | | |
| Edit Details | ""Tying the Knot"" Pocket Notes White Full Grain Leather" | 36 | | | 11358494 | GiGi New York / Graphic Image | handbags & accessories | qjo 2334 | usd | true | | PL5-KNOT-WHT | | |
| Edit Details | "\$100 LEGAL SEA FOODS GIFT CARD" | 100 | | | 10407384 | Legal Sea Foods Gourmet Gift | | qjo 1161 | USD | true | | "RCD100" | | |

Figure 38: Products Screen

Admin Product Edit Screen

This screen allows admin to perform edit operation over a specific product.

The screenshot shows the 'productEdit' screen. On the left is a category sidebar with 'Women' selected. The main area contains a form with the following fields:

| Field | Value |
|--------------------|--|
| AdID | 11350107 |
| AdvertiserID | 3926812 |
| AdvertiserCategory | Stockings > Seamed Stockings |
| BuyUrl | http://www.tkgqhce.com/click-3680720-11350107-1471190845896?url=http%3A%2F%2Fthetightspot.com%2Fdreamgirl-sheer-seamed-stockings&cjsku=DG022 |
| CatalogID | cjo.3059 |
| Currency | EUR |
| Description | " Dreamgirl Sheer Seamed Stockings" |
| ImageUrl | http://thetightspot.com/media/catalog/product/d/d/dreamgirl-sheer-seamed-stockings-bfb_1.jpg |
| InStock | true |
| ManufacturerSKU | |
| Name | " Dreamgirl Sheer Seamed Stockings" |

Figure 39: Products Edit Screen

Admin Product Detail Screen

This screen allows admin to see details of products

The screenshot shows the 'productDetails' screen. On the left is a category sidebar with 'Women' selected. The main area displays the following product details:

| Field | Value |
|--------------------|--|
| AdID | 11350107 |
| Advertiser | The Tight Spot |
| AdvertiserCategory | Stockings > Seamed Stockings |
| BuyUrl | http://www.tkgqhce.com/click-3680720-11350107-1471190845896?url=http%3A%2F%2Fthetightspot.com%2Fdreamgirl-sheer-seamed-stockings&cjsku=DG022-C |
| CatalogID | cjo.3059 |
| Currency | EUR |
| Description | " Dreamgirl Sheer Seamed Stockings" |
| ImageUrl | http://thetightspot.com/media/catalog/product/d/d/dreamgirl-sheer-seamed-stockings-bfb_1.jpg |
| InStock | true |
| ManufacturerSKU | |
| Name | " Dreamgirl Sheer Seamed Stockings" |
| Price | 5.82 |
| RetailPrice | |
| SalePrice | |
| SKU | DG022-C |
| UPC | |
| ISBN | |
| ProductAutoKey | 85010 |
| CreateDate | 18-Sep-16 12:00:00 AM |
| ModifyDate | 18-Sep-16 12:00:00 AM |
| ManufacturerName | |

At the bottom left, there are links: [Edit](#) | [Back to List](#)

Figure 40: Products Detail Screen

Admin offers Screen

This Screen allows admin to see all the offers and select a particular offer to see its details, edit it or perform analysis.

Admin Best Offers

Advertisers-

Offers-

Products-

Hello ahmad@gmail.com!

Log off

Product

Search

Search

Category

Women

Computer Software

Consumer Electronics

Hotel

Gifts

Vacations

Virtual Malls

Jewelry

Apparel

Parts and Accessories

Web Hosting/Servers

View Offers

| Advertiser | Category | Language | LinkID | LinkType | Promotion | CouponCode | RelationStatus | SalesCommission | SevenDayEPC | ThreeMonthEPC | LinkName |
|--|---------------|----------|--------|----------|-----------|------------|----------------|-----------------|-------------|---------------|-------------------------------|
| Edit Details Add | Online Sports | Sports | en | 11118673 | Banner | | joined | 10.00% | N/A | N/A | 180x150 College Football 2012 |
| Edit Details Add | Online Sports | Sports | en | 11118675 | Banner | | joined | 10.00% | N/A | N/A | 234x60 College Football 2012 |
| Edit Details Add | Online Sports | Sports | en | 11118683 | Banner | | joined | 10.00% | N/A | N/A | 468x60 College Football 2012 |
| Edit Details Add | Online Sports | Sports | en | 11118680 | Banner | | joined | 10.00% | N/A | N/A | 300x250 College Football 2012 |
| Edit Details Add | Online Sports | Sports | en | 11118685 | Banner | | joined | 10.00% | N/A | N/A | 768x90 College Football 2012 |
| Edit Details Add | Online Sports | Sports | en | 11144282 | Banner | | joined | 10.00% | N/A | N/A | 160x600 NFL Football |

Figure 41: Offers Screen

Admin offer detail

This page allows admin to see details of offer

Admin Best Offers

Advertisers -

Offers -

Products -

Hello ahmad@gmail.com!

Log off

Product >

Search

Search

Category

Women

Computer Software

Consumer Electronics

Hotel

Gifts

Vacations

Virtual Malls

Jewelry

Apparel

Parts and Accessories

Web Hosting/Servers

Malls

Offer Details

| | |
|----------------------|--|
| AdvertiserID | 237343 |
| CategoryID | 112 |
| ClickCommission | 0 |
| LanguageID | 9 |
| LeadCommission | 3 |
| Description | 24" AOC e2425Swd - LED monitor E2425SSWD 9830665 |
| Destination | http://www.pcm.com/p.pdf?dpno=9830665&store=pcm&source=PWBCJTEXT |
| LinkID | 12573633 |
| LinkIdTypeID | 3 |
| PerformanceIncentive | false |
| PromotionEndDate | 30-Sep-16 8:59:00 AM |
| PromotionStartDate | 17-Mar-16 9:00:00 AM |
| PromotionID | 41 |
| CouponCode | |
| RelationStatus | joined |
| SalesCommission | 1.00% - 3.00% |
| SevenDayEPC | N/A |
| ThreeMonthEPC | N/A |
| ClickUrl | http://www.tkgqhc.com/click-3680720-12573633-1463589714000 |
| LinkName | 24" AOC e2425Swd - LED monitor E2425SSVD 9830665 |
| CreateDate | 19-Sep-16 12:00:00 AM |
| modifyDate | 19-Sep-16 12:00:00 AM |
| Height: | 0; |
| Width: | 0; |
| JavaScriptCode | <form name="CJ12573633X1" method="POST" style="margin:0px;display:inline" action="http://www.jdoquery.com/click"><input type="hidden" name="aid" value="12573633"><input type="hidden" name="pid" value="3680720"><input type="hidden" name="lastUpdatedDate" value="1463589714000">Save \$90! 24" AOC e2425Swd - LED monitor only \$109.99 (was \$199.99) </form>;Save \$90! 24" AOC e2425Swd - LED monitor only \$109.99 (was \$199.99) </img |
| HTMLCode | |

Figure 42: Offer Detail Screen

Offer Edit

This page allows admin to edit offers

The screenshot shows the 'Offer Edit' interface. On the left is a 'Category' sidebar with a list of categories: Women, Computer Software, Consumer Electronics, Hotel, Gifts, Vacations, Virtual Malls, Jewelry, Apparel, Parts and Accessories, Web Hosting/Servers, and Malls. The main area is titled 'Offer Edit' and contains a form with the following fields: AdvertiserID (79362), CategoryID (Health and Wellness), ClickCommission (0), LanguageID (9), LeadCommission, Description (Proven Drug Testing Solutions), Destination (http://www.testdear.com), LinkID (35602), LinkTypeID (3), PerformanceIncentive (true), PromotionEndDate (01-Jan-01 12:00:00 AM), and PromotionStartDate. The top navigation bar includes 'Admin Best Offers', 'Advertisers', 'Offers', and 'Products', along with a user profile 'Hello ahmad@gmail.com!', a 'Log off' link, and a search bar.

Figure 43: Offer Edit Screen

Offer Add

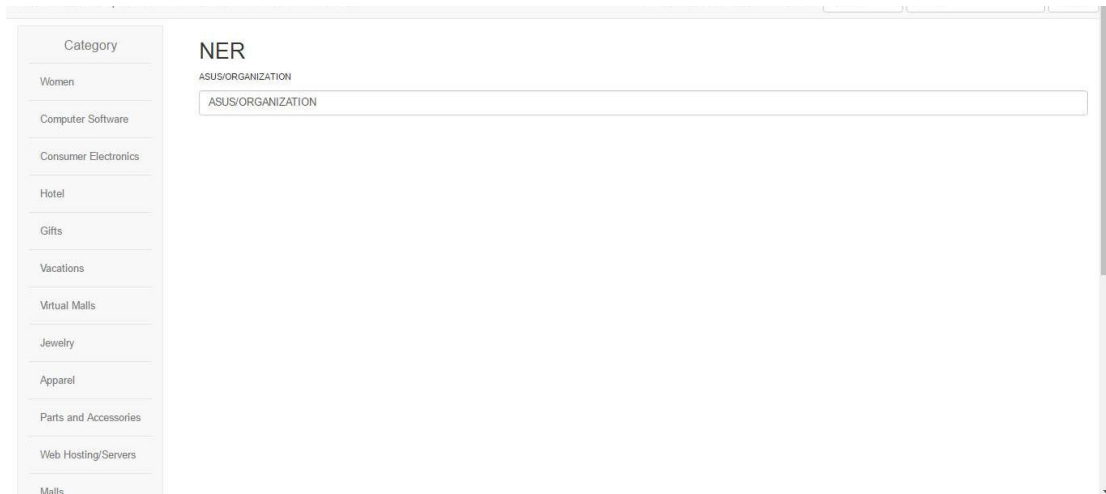
This module performs analysis on the description of offer and show extracted information in the fields of form moreover user can also add its own details

The screenshot shows the 'Add Offer' interface. On the left is the same 'Category' sidebar as in Figure 43. The main area is titled 'Add Offer' and displays the product description 'PNY Geforce Gt610 PCIe 2.0 1GB DDR3 VCGGT610XPB 9204676'. Below this is a form with fields for: Save Amount, Off Amount, End Date, Eligible Amount, Site Wide, Free Shipping, and a code field. An 'Add' button is located at the bottom right of the form. At the bottom left, there are links for 'NER Analysis' and 'Back to List'. The top navigation bar is identical to the one in Figure 43.

Figure 44: Add Offers Screen

NLP NER Analysis

It allows admin to perform NER analysis over description of offer

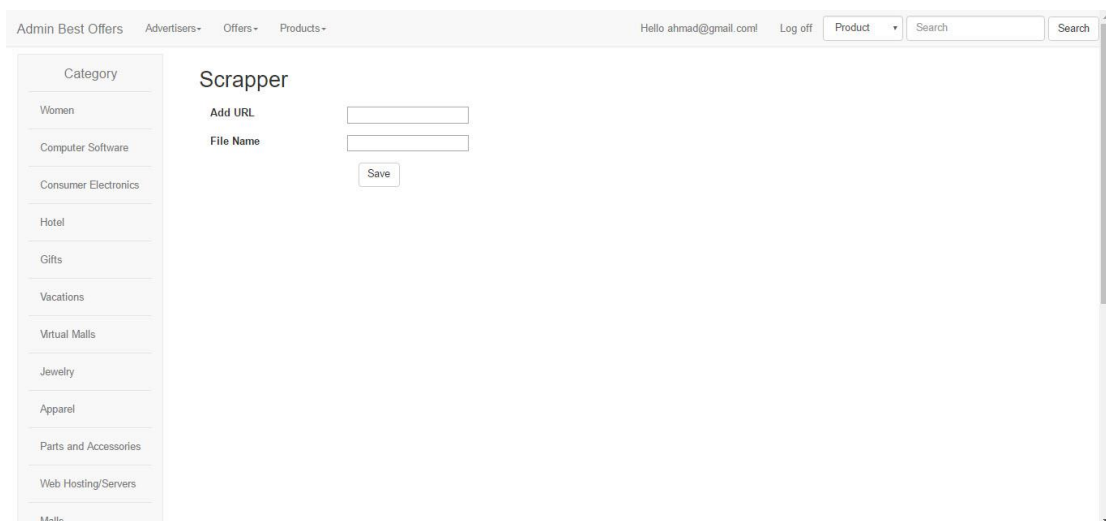


The screenshot shows a web application interface for NER analysis. On the left, there is a vertical sidebar with a 'Category' dropdown menu. The menu is open, showing a list of categories: Women, Computer Software, Consumer Electronics, Hotel, Gifts, Vacations, Virtual Malls, Jewelry, Apparel, Parts and Accessories, Web Hosting/Servers, and Malls. The main content area is titled 'NER' and contains a text input field with the placeholder text 'ASUS/ORGANIZATION'. Below the input field, there is a small text label 'ASUS/ORGANIZATION'.

Figure 45: NER Analysis

Scrapping Screen

This screen allows admin to scrap mentioned site



The screenshot shows a web application interface for a scrapping tool. At the top, there is a navigation bar with links: Admin Best Offers, Advertisers-, Offers-, and Products-. On the right side of the navigation bar, there is a user profile section with the text 'Hello ahmad@gmail.com!', a 'Log off' link, and a 'Product' dropdown menu. Below the navigation bar, there is a search bar with the text 'Search' and a 'Search' button. The main content area is titled 'Scrapping' and contains two input fields: 'Add URL' and 'File Name'. Below these fields, there is a 'Save' button. On the left, there is a vertical sidebar with a 'Category' dropdown menu, which is open, showing a list of categories: Women, Computer Software, Consumer Electronics, Hotel, Gifts, Vacations, Virtual Malls, Jewelry, Apparel, Parts and Accessories, Web Hosting/Servers, and Malls.

Figure 46: Scrapping

Keyword Cloud

This screen allows admin to see top keywords

| Category | Keyword Cloud | |
|----------------------|---------------|-------|
| Women | Keyword | Count |
| Computer Software | com | 9917 |
| Consumer Electronics | free | 7607 |
| Hotel | off | 7101 |
| Gifts | shipping | 5218 |
| Vacations | shop | 5182 |
| Virtual Malls | save | 4207 |
| Jewelry | link | 3902 |
| Apparel | new | 3451 |
| | banner | 3422 |
| | sale | 3251 |
| | code | 2744 |
| | online | 2291 |

Figure 47: Key Cloud

EVALUATION

Test Cases

Admin Login Test Case 1

Table 30: Test Case Login

| | | | |
|-------------------------|--|-----------------------|--------------|
| Test Case ID: | 1 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin Login | Date: | |
| Purpose: | To check Login component | | |
| Pre-Condition | Admin should not be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Navigate to main Page2. Click on Login3. Enter Correct Password | | |
| Expected Result: | Admin should successfully Login | | |
| Result: | Admin Login successfully | | |
| Test Case State | Pass | | |
| Post Condition | Admin got login | | |

Admin Login Test Case 2

Table 31: Test Case Login 2

| | | | |
|-------------------------|--|-----------------------|--------------|
| Test Case ID: | 2 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin Login | Date: | |
| Purpose: | To check Login component | | |
| Pre-Condition | Admin should not be in login state | | |
| Steps to Perform | 1. Navigate to main Page 2. Click on Login 3. Enter Wrong Password | | |
| Expected Result: | Login should fail | | |
| Result: | Login failed | | |
| Test Case State | Pass | | |
| Post Condition | No login | | |

Admin Rights Test Case

Table 32: Test Case Admin Rights

| | | | |
|-----------------------|--|-----------------------|--------------|
| Test Case ID: | 3 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin Rights | Date: | |
| Purpose: | To check whether rights have been properly assigned | | |
| Pre-Condition | Admin should not be in login state | | |
| Steps to | 1. Enter admin side address in address bar like site/admin | | |

| | |
|-------------------------|--------------------------|
| Perform | |
| Expected Result: | Redirect to login page |
| Result: | Redirected to login Page |
| Test Case State | Pass |
| Post Condition | No change |

Admin editing advertisers test case

Table 33:Test Case Edit Advertiser

| | | | |
|-------------------------|---|-----------------------|--------------|
| Test Case ID: | 4 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin editing advertisers | Date: | |
| Purpose: | To check whether admin can properly edit advertisers | | |
| Pre-Condition | Admin should be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Go to view advisers3. Click on edit advertisers4. Change the information | | |
| Expected Result: | Advertiser information should get updated | | |
| Result: | Advertiser information got updated | | |
| Test Case State | Pass | | |
| Post Condition | Updated advertiser information got saved in database | | |

Admin editing products test case

Table 34: Test Case Edit Product

| | | | |
|-------------------------|--|-----------------------|--------------|
| Test Case ID: | 5 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin editing products | Date: | |
| Purpose: | To check whether admin can properly edit products | | |
| Pre-Condition | Admin should be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Go to view products3. Click on edit products4. Change the information | | |
| Expected Result: | Products information should get updated | | |
| Result: | Products information got updated | | |
| Test Case State | Pass | | |
| Post Condition | Updated Product information got saved in database | | |

Admin editing offers

Table 35: Test Case Edit Offers

| | | | |
|-----------------------|---|-----------------------|--------------|
| Test Case ID: | 6 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin editing offers | Date: | |
| Purpose: | To check whether admin can properly edit offers | | |

| | |
|-------------------------|---|
| Pre-Condition | Admin should be in login state |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Go to view advisers3. Click on edit advertisers4. Change the information |
| Expected Result: | Advertiser information should get updated |
| Result: | Advertiser information got updated |
| Test Case State | Pass |
| Post Condition | Updated advertiser information got saved in database |

Admin performing keyword analysis over offers test case

Table 36: Test Case Key Analysis

| | | | |
|-------------------------|---|-----------------------|--------------|
| Test Case ID: | 7 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin performing keyword analysis over offers | Date: | |
| Purpose: | To check whether keyword analysis is working or not | | |
| Pre-Condition | Admin should be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Go to view offers3. Click on add offers | | |
| Expected Result: | Admin should see the extracted information in form | | |
| Result: | Admin see the extracted information in form | | |
| Test Case State | Pass | | |

| | |
|-----------------------|--|
| Post Condition | |
|-----------------------|--|

Admin performing NER/NLP analysis over offers test case

Table 37:Test Case NER

| | | | |
|-------------------------|---|-----------------------|--------------|
| Test Case ID: | 8 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin performing NER/NLP analysis over offers | Date: | |
| Purpose: | To check whether NLP analysis is working or not | | |
| Pre-Condition | Admin should be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Go to view offers3. Click on add offers4. Click on NER | | |
| Expected Result: | Admin should see the information got through NLP | | |
| Result: | Admin see the information got through NLP | | |
| Test Case State | Pass | | |
| Post Condition | | | |

Admin adding offers test case

Table 38:Test Case Admin Adding Offers

| | | | |
|----------------------|---------------------|-----------------------|--------------|
| Test Case ID: | 9 | System Tester: | Ahmad Saleem |
| Test Case | Admin adding offers | Date: | |

| | | | |
|-------------------------|--|--|--|
| Name | | | |
| Use Case Ref: | | | |
| Purpose: | To check whether admin can successfully approve/add offers | | |
| Pre-Condition | Admin should be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Go to view offers3. Click on add offers4. Click on add button in form | | |
| Expected Result: | Updated information should get saved in database | | |
| Result: | Updated information got saved in database | | |
| Test Case State | Pass | | |
| Post Condition | Offer state is updated | | |

Approved offers information test case

Table 39: Test Case Approved Offers

| | | | |
|-------------------------|--|-----------------------|--------------|
| Test Case ID: | 10 | System Tester: | Ahmad Saleem |
| Test Case Name | Approved offers information | Date: | |
| Purpose: | Admin getting information whether an offer has been approved or not | | |
| Pre-Condition | Admin should be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Go to view offers3. Look for offer that has been added4. Click on details | | |

| | |
|-------------------------|--|
| | |
| Expected Result: | A box displaying that offer has been added should be on page |
| Result: | A box displaying that offer has been added appeared on page |
| Test Case State | Pass |
| Post Condition | Offer state is updated |

Admin scrapping website test case

Table 40:Test Case Scrapping

| | | | |
|-------------------------|--|-----------------------|--------------|
| Test Case ID: | 11 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin scrapping website | Date: | |
| Purpose: | To check whether admin can successfully scrape a website or not | | |
| Pre-Condition | Admin should be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Click on scrape website3. Enter Site name4. Enter file name5. Click on scrape button | | |
| Expected Result: | Page should show a success message and all text data of that file should get saved | | |
| Result: | Page showed a success message and all text data of that file got saved | | |
| Test Case State | Pass | | |
| Post Condition | Admin successfully scraped website | | |

Admin getting top keywords test case

Table 41:Test Case Top Key Words

| | | | |
|-------------------------|---|-----------------------|--------------|
| Test Case ID: | 12 | System Tester: | Ahmad Saleem |
| Test Case Name | Admin seeing top keywords | Date: | |
| Purpose: | To check whether admin can check top keywords or not | | |
| Pre-Condition | Admin should be in login state | | |
| Steps to Perform | <ol style="list-style-type: none">1. Go to admin side2. Click on Keyword Cloud | | |
| Expected Result: | Page showing top keywords in a database should appear | | |
| Result: | Page showing top keywords in a database got appeared | | |
| Test Case State | Pass | | |
| Post Condition | Admin successfully checked top keywords | | |

User searching products test case

Table 42: Searching Products

| | | | |
|-------------------------|--|-----------------------|--------------|
| Test Case ID: | 13 | System Tester: | Ahmad Saleem |
| Test Case Name | User Searching products | Date: | |
| Purpose: | To check whether user can search products or not | | |
| Pre-Condition | | | |
| Steps to Perform | <ol style="list-style-type: none">1. Select products from a dropdown in menu bar2. Enter product name3. Click on search button | | |

| | |
|-------------------------|---|
| Expected Result: | Page showing matched products should appear |
| Result: | Page showing matched products appeared |
| Test Case State | Pass |
| Post Condition | User successfully searched products |

Particular category offers test case

Table 43: Test Case Category Offer

| | | | |
|-------------------------|---|-----------------------|--------------|
| Test Case ID: | 14 | System Tester: | Ahmad Saleem |
| Test Case Name | Particular category offers | Date: | |
| Purpose: | User should be able to check offers by a category | | |
| Pre-Condition | | | |
| Steps to Perform | 1. Select category from sidebar 2. Select offers | | |
| Expected Result: | Page showing matched offers should appear | | |
| Result: | Page showing matched offers appeared | | |
| Test Case State | Pass | | |
| Post Condition | User successfully located offers of a particular category | | |

Particular advisers offer test case

Table 44: Particular Advertiser

| | | | |
|----------------------|----|-----------------------|--------------|
| Test Case ID: | 15 | System Tester: | Ahmad Saleem |
|----------------------|----|-----------------------|--------------|

| | | | |
|-------------------------|---|--------------|--|
| Test Case Name | Particular advisers offer | Date: | |
| Use Case Ref: | | | |
| Purpose: | User should be able to check offers by advertiser | | |
| Pre-Condition | | | |
| Steps to Perform | <ol style="list-style-type: none">1. Select top advertisers2. Select a particular advertiser3. Select view offers | | |
| Expected Result: | Page showing matched offers should appear | | |
| Result: | Page showing matched offers appeared | | |
| Test Case State | Pass | | |
| Post Condition | User successfully located offers by a particular advertiser | | |

Results

The Project was successfully tested under controlled environment according to the above mentioned test cases. Some more test cases were also used to test the project but are not mentioned here because the level of detail would get too high. All the test cases returned expected results and small bugs that were discovered during this process were fixed.

Conclusions & Future Work

Internet is full of online stores so we decided to make a site that gathers data from these sellers and show it on our site so that a user can get a lot on one site and to improve its search ranking and data management we had to find some way to implement schema.org systematically.

At the start of our project we had no idea what to do. So we did a lot of research from google search algorithms to Natural language processing libraries it not only helped us in making our project but increased our knowledge base to a great extent. Currently the system that we have developed is showing good performance and is completely implementable.

There are some areas that can still be improved and I will put them in future work few of them are related to site and some are related to information extraction

- Data capturing module has to be made more dynamic and it has to be embedded inside website.
- Some more facilities could be provided on user or admin side like search by date and user review
- Keyword extraction technique could be improved more by recognizing more patterns.
- NLP is to be made more effective by training its annotators
- A hybrid approach could be used bringing in both nlp and keyword analysis to improve information extraction
- Scrapping module was also made with an intention to extract missing information. So we could extract missing information about an offer or product from its parent site. This module could be mixed along hybrid analysis approach or some special approach could be designed for this module that could make it dynamic.

References

- [1] *Form based schema implementation*. (n.d.). Retrieved from <http://www.microdatagenerator.com/MicrodataGenerator/generator.html>
- [2] *Good Relations*. (n.d.). Retrieved from <http://www.heppnetz.de/ontologies/goodrelations/v1#Offering>
- [3] *Google Algorithms*. (n.d.). Retrieved from <https://moz.com/blog/google-algorithm-cheat-sheet-panda-penguin-hummingbird>
- [4] *Google Panda*. (n.d.). Retrieved from <http://www.seo-theory.com/2013/04/10/how-the-google-panda-algorithm-works/>
- [5] *panda quality check*. (n.d.). Retrieved from <https://webmasters.googleblog.com/2011/05/more-guidance-on-building-high-quality.html>
- [6] *RankBrain*. (n.d.). Retrieved from <http://searchengineland.com/library/google/google-rankbrain>
- [7] *Schema.org*. (n.d.). Retrieved from <http://schema.org/>
- [8] *The RDFa Content Editor - From WYSIWYG to WYSIWYM*. (n.d.). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6340208
- [9] Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S. and McClosky, D., 2014, June. The Stanford CoreNLP Natural Language Processing Toolkit. In *ACL (System Demonstrations)*
- [10] Ronallo, J., 2012. HTML5 Microdata and Schema. org. Code4Lib Journal, 16
Khalili, A. and Auer, S., 2013, October. Wysiwy authoring of structured content based on schema. org. In *International Conference on Web Information Systems Engineering* (pp. 425-438). Springer Berlin Heidelberg
- [11] Patel-Schneider, P.F., 2014, October. Analyzing schema. org. In *International Semantic Web Conference* (pp. 261-276). Springer International Publishing.
- [12] Robert Meusel, Christian Bizer and Heiko Paulheim, "A Web-scale Study of the Adoption and Evolution of the schema.org Vocabulary over Time", 2015.
- [13] AnHai Doan, Jeffrey F. Naughton, Akanksha Baid, Xiaoyong Chai, Fei Chen, Ting Chen, Eric Chu, Pedro DeRose, Byron Gao, Chaitanya Gokhale, Jiansheng Huang, Warren Shen, Ba-Quy Vuong, "The Case for a Structured Approach to Managing Unstructured Data".
- [14] Harish Jadhao, Dr. Jagannath Aghav, Anil Vegiraju, "Semantic Tool for Analysing Unstructured Data", 2012.
- [15] I. Hickson. HTML Microdata. <http://www.w3.org/TR/microdata/>, 2011.
<http://www.datacaptureexperts.com.au/how-to-win-unstructured-data-management-challenge/>
- [16] O'Neill S, Curran K. "The Core Aspects of Search Engine Optimization Necessary to Move up the Ranking". *International Journal of Ambient Computing and Intelligence*. 2011; 3(4):62-70.