Information hiding in Linked Opened Data

Asaad Sabah Hadi
Babylon University, Information Technology College, Software Department.
asaadsabah@yahoo.com

Abstract

Information hiding is very important for securing an information from the malicious attackers. Steganography is the science of hiding the original information (message) into a carrier. The user of the internet usually need to store, send and receive a private information. One of the most famous method for doing this is the encryption, which it converts the information into another format that is not understood to attackers. The major drawback of the encryption is that it cannot hide the existence of the data, because the result is still data so if there is enough time someone may decrypt the encrypted data. The steganography is the solution for this problem. The Linked data are the idea for linking different web data sources by establishing a link between them. The link describes the relationships between all these sources. The main idea about this paper is securing the important information inside Linked opened data . Usually, for embedding secret information into a cover message we need redundant data or noise, therefore, we use the least important features in the RDF to hide our important information. In order to increase the security of our important information we encrypt it using the Beaufort cipher. We implement the proposed method using two RDF representation for Alzheimer disease and Amyloid precursor protein and give accepted security for our important information.
Keywords : Steganography, LOD, RDF, information hiding,Cipher, Beaufort Cipher.

الخلاصة

يعتبر اخفاء المعلومات مهم جدا لتأمين المعلومات من الهجمات الخبيثة. من الممكن تعريف اخفاء البيانات على انها عملية اخفاء المعلومات الاصلية (الرسالة) في الناقل. يحتاج مستخدم الانترنت عادة الى تخزين وارسال واستقبال المعلومات الخاصة به وتعتبر تقنية التشفير واحدة من الطرق المشهورة للقيام بذلك، حيث يقوم بتحويل المعلومات الى صيغة اخرى غير مفهومة للمهاجمين. العيب الرئيسي للتشفير هو عدم امكانية اخفاء البيانات وذلك لان الناتج من التشفير يبقى معلومات ايضا وفي حالة توفر الوقت يستطيع الشخص فك التشفير. تعتبر اخفاء البيانات واحدة من الحلول لهذه المشكلة. يعتبر مفهوم البيانات المترابطة من المفاهيم الحديثة لربط مصادر البيانات على شبكة الانترنت من خلال خلق ترابط بينهم. يوضح الترابط نوعية العلاقة بين تلك المصادر. الفكرة الاساسية في هذا البحث هو تأمين المعلومات الهامة في داخل (Linked Opened Data). عادة ، ولغرض اخفاء البيانات المهمة في غطاء معين سنحتاج الى دراسة هذا الغطاء واستخراج البيانات المكررة والضوضاء لغرض الاستفادة منها في عملية الاخفاء ولذلك فاننا قمنا باختيار الخصائص الاقل اهمية في (RDF) لغرض اخفاء المعلومات المهمة فيها. لغرض زيادة سرية البيانات قمنا باستخدام التشفير بواسطة طريقة (Beaufort) وتم تطبيق الطريقة المقترحة على نوعين من التمثيل واثبتت نجاحها في اخفاء المعلومات المهمة.

الكلمات المفتاحية : اخفاء المعلومات، ربط فتح البيانات , اطار وصف الموارد، اخفاء البيانات، التشفير ، تشفير بيوفورت.

I. Introduction

The information in the web can be accessed by human and machine where the web can be described as a space of an information [Bo Leuf 2006]. The Semantic Web is not a new web but is an extension to the current web in a way that make incorporations between computer and people to create a universal medium to exchange the data [Andre Freitas 2012][Bo Leuf 2006]. Because all the information in the web is accessible by everyone, therefore, there is a need to make access control to the important information in the web [Christian Bizer 2009]. There are two important approaches for specifying that are allowed to use which data [Richard Sheffield 2009]. The first is to use cryptographic techniques to protect the sensitive data. The second is to make access rights in order to control the data access and secure the communication channels. The RDF is a special language that represents the

information about the World Wide Web resources and was intending for representing the web resources metadata[Eric Miller 1998][Edgard Marx 2012].

The common techniques for encrypting sensitive data in the RDF are cutting the data from the original graph and storing it in a separate file, then encrypt the file and link it to the original RDF. The drawback for this techniques is : (a) there are no rules for re-integrating the data to the RDF after decryption (b) the encrypted files are not RDF-compliant (c) different physical resources for the original RDF. (d) the linking is done manually. In this paper we will hide the secret information in the same RDF graph.

II. Linked Opened Data

The linked data represent the objects and relation between them. The linked data web contains data in multiple areas and this area grows exponentially. The representation of these data is very important in order to facilitate information retrieval [Andre Freitas 2012]. The Hypertext links allow the user to traverse the World Wide Web information using the web browser[Christian Bizer 2009]. In the last years the web has formularized from a very global information space of linked documents into one that data and documents are linked[Christian Bizer 2009]. The Web of data gives rise to new types of application and opens new possibilities for domain-specific applications[Richard Sheffield 2009]. The Opened Link Data use the Web to create type links between data from different sources. The primary units for the hypertext Web are HTML (HyperText Markup Language) documents connected by hyperlinks[Eric Miller 1998,Christian Bizer 2009]. The Linked Data lean on documents that contain data in RDF (Resource Description Framework) format. The Linked Data principles can be summarized as [Christian Bizer 2009][Michael Hausenblas 2009] : (1) Use URIs (Universal Resource Identifiers) as names for the things. (2) Use the HTTP(Hypertext Transfer Protocol) URIs in order to make people look up for those names.(3) Use RDF to provide useful information. (4) Include links to other URIs to discover more things.

III. RDF

RDF(Resource Description Framework) is the infrastructure that facilitates the exchange, encoding and reuse of the metadata[Eric Miller 1998]. It provides the publishing for both human and machine readable vocabularies. RDF uses XML (eXtesible Markup Language) as a syntax for exchanging and processing of metadata[Edgard Marx 2012]. RDF gives a model for describing the resource. The resource has properties (attributes). It defines the resource for any object that is uniquely identified by a URI(Uniform Resource Identifier) [Eric Miller 1998]. RDF contains statements that represent the information in the form of : subject, predicate and object.

IV. Semantic Web

The Semantic web provides an infrastructure that enables web pages and services, database and programs to the web[Bo Leuf 2006]. It is an extension for the current web that gives the meaning of the information in a form of vocabularies understood by both computers and humans[Christian Bizer,Jens Lehmann 2009][Christian Bizer,Tom Heath 2009].

The classical web lacks the existence of a semantic structure, therefore it is very difficult for the computer to understand the information provided by a user[Michael Hausenblas 2010].

The classical web suffer from many problems [Bo Leuf 2006][Christian Bizer,Jens Lehmann 2009] [Michael Hausenblas 2009]:

- Lack of universal format for representing the information in the web pages.
- There is an ambiguity of the information due to the poor interconnection between them.
- Lack of web content structure.

Tim Burner Lee is the foundation father of the Semantic Web and it gives many benefits like Web-Base, Extensibility, Ability of domain-driven models to be interlinked, model expressiveness, use of standard language[Bo Leuf 2006][Christian Bizer, Jens Lehmann 2009].

V. Information Hiding

The security requirements involve hiding the important information from an outside observer[Dominic Hughes 2004].

VI. Steganography & Cryptography

In the digital world, cryptography and steganography are used to protect information from attackers. They are both work good alone , but they work excellent together[Greg kipper, 2004][Arvind Kumar 2010][Bo Leuf 2006].

The term Steganography means " Cover writing" whereas cryptography means "secret writing". Figure (1) shows the Steganography whereas Figure (2) illustrates Cryptography[Arvind Kumar 2010][Bo Leuf 2006].

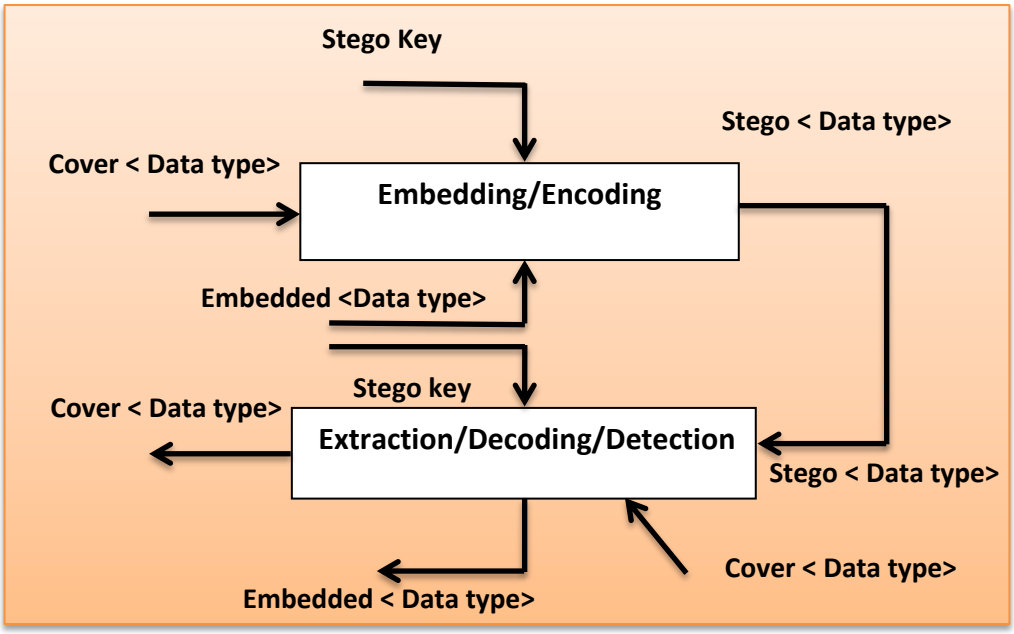

Figure (1) Steganography

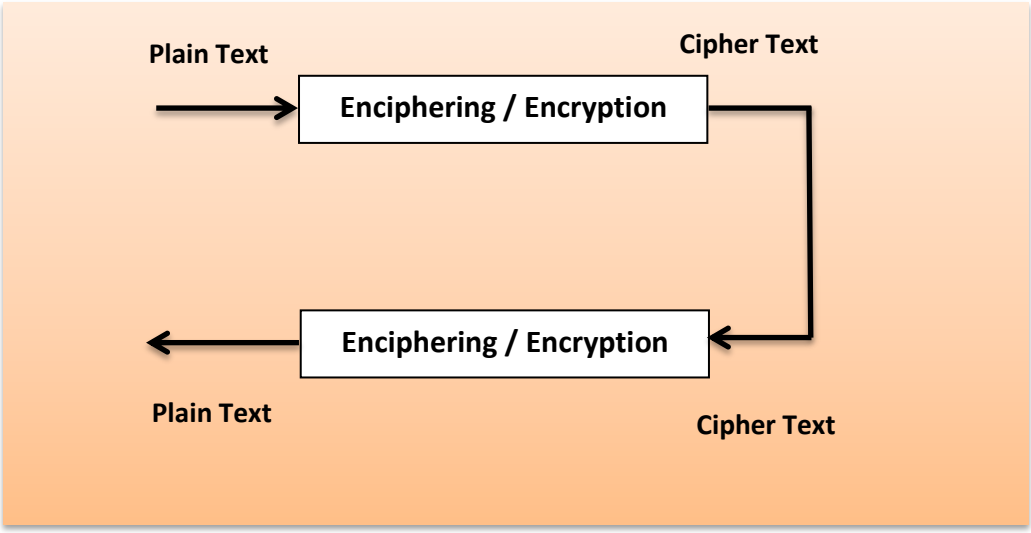

Figure (2) Cryptography

Logically, there are three types of steganography [Bret Dunbar 2002] [Bo Leuf 2006]:

- Pure steganography, which does not require exchanging cipher key.
- Secret key steganography, that requires exchanging of secret key prior to communication.
- Public key Steganography, that uses public key and private key in order to secure the communication. It uses the same concept of public key cryptography.

VII. Cipher System

The Classification of Cipher system can be shown in Figure (4).In our paper , we used the beaufort cipher which is a polyalphabetic substitution cipher. In Beaufort cipher we have a table (like Vigenere cipher table) as shown in Figure (3) below, and our secret information is named plaintext and there is a keyword (in this paper the keyword is the keyword_statement) that is duplicated according to the plaintext length. After that we find the cipher text by using the intersection between plaintext character with keyword character according to the vigenere table.

Ex. Plaintext = my computer is Toshiba
Keyword = bestbestbestbestbestbe
Cipher text = pgqfppyaxnkbiqamtds

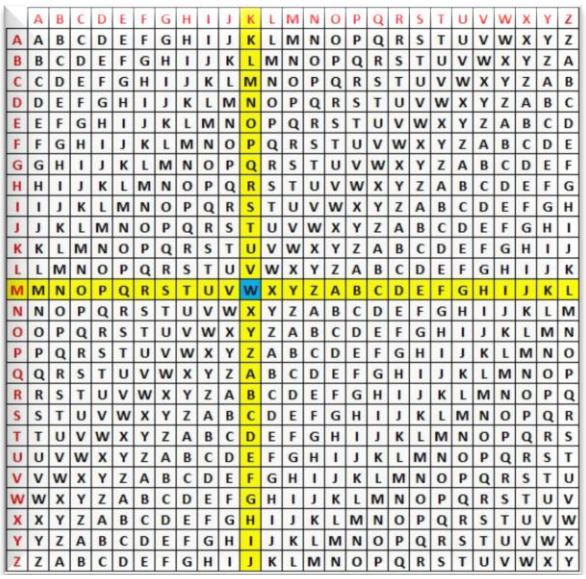

Figure (3) Vigenere table

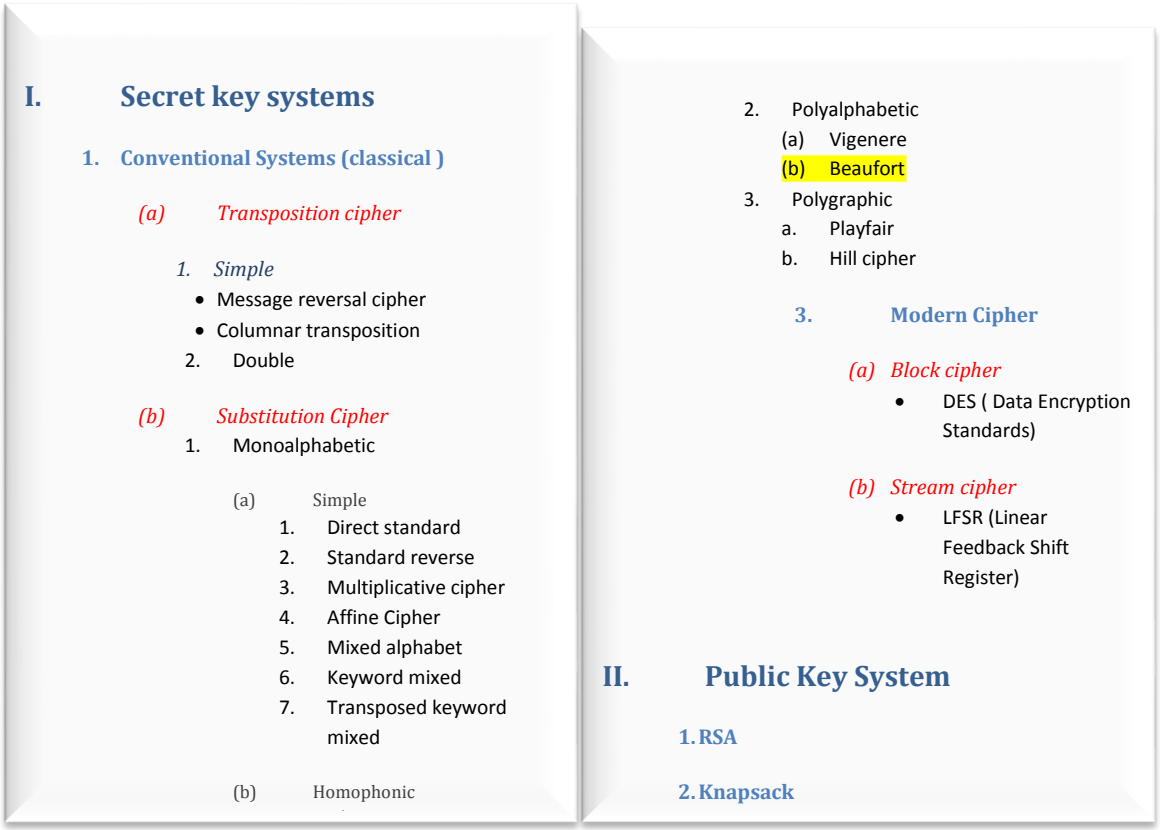

Figure (4) Classification of cipher system

VIII. Framework Architecture

In the beginning , the proposed system reads the RDF representation of the cover in the form of N-Triples(It Can convert the RDF/XML into N-triples) and analyzes it into a number of statements (Triple statements). After that we sort the statements in increasing order according to their importance , then we take the first rank (the lowest important statement) and name it keyword_predicate and name

the statement that contains it as keyword_statement. After that we use this statement to encrypt our important information by using Beaufort cipher .Figure (5) below shows a simple framework for our system :

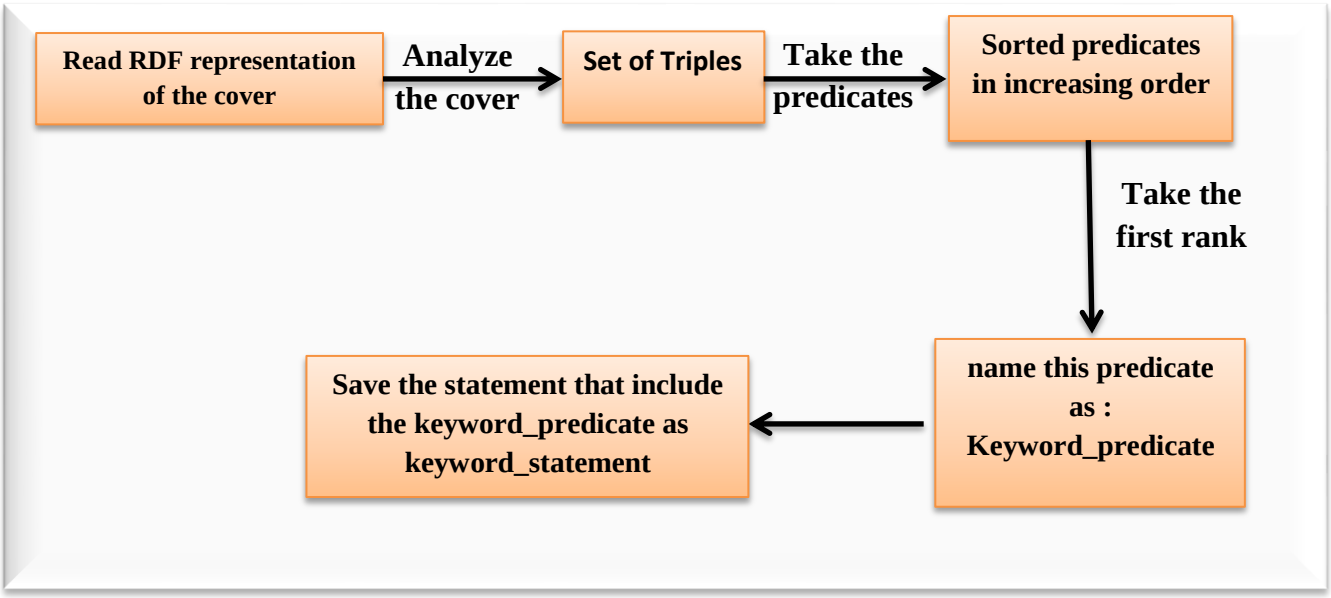

Figure (5) –a Specify the proper cover

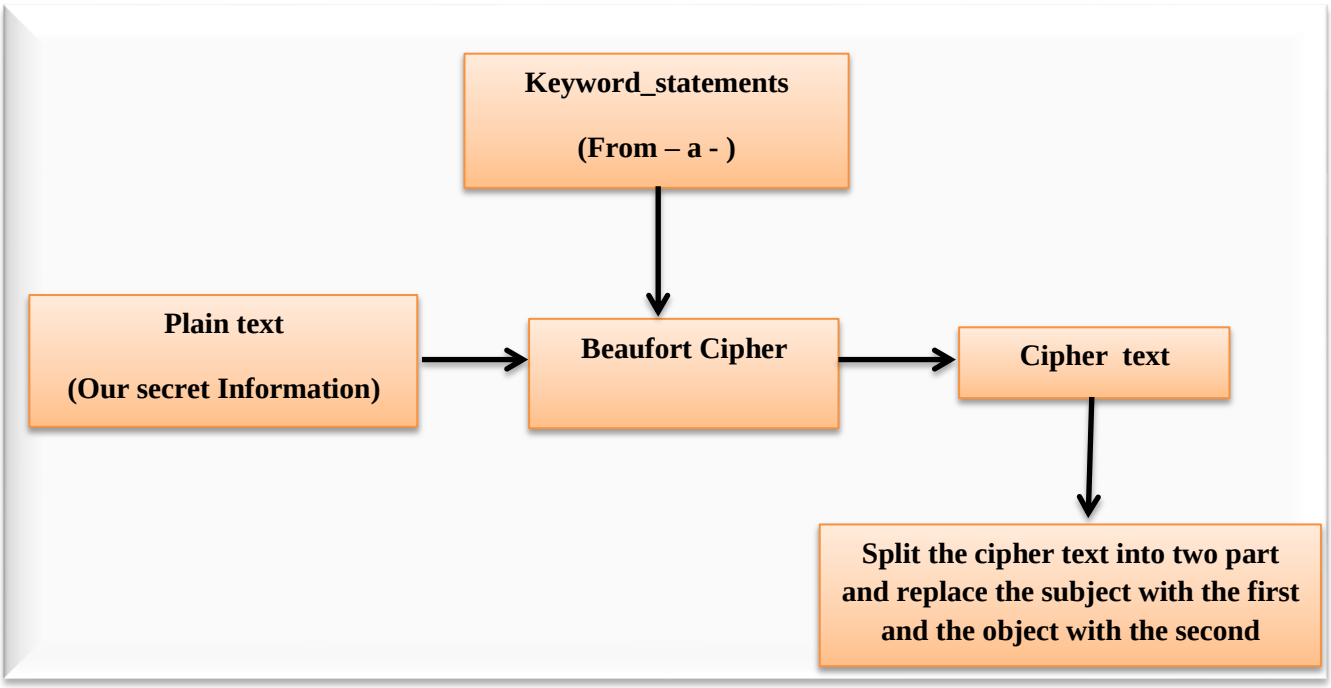

Figure (5) –b Ciphering Schema

In order to test our proposed system, we take some samples :-

1. Alzheimer disease

This RDF file contains (66) Triples : the no. of subject is (1), the number of predicate is (39), the number of object is (59).

Our secret message is ' **our army begins the attack after one hour**'. After analyzing the RDF file we take this statement as keyword statement : 'Alzhemers discease subject category Aphasias' , where the subject is (Alzhemers

discease), the predicate is (subject), the object is (category Aphasias) which is shown in Figure (6) below.

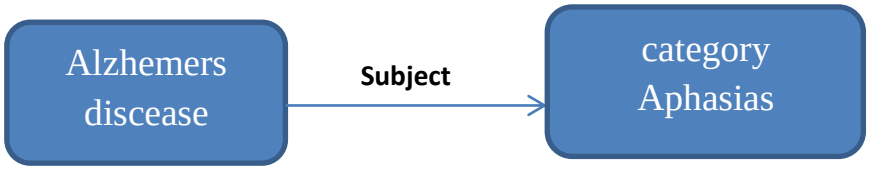

Figure (6) Keyword_statement for Alzhemers

Plaintext = our army begin the attack after one hour
Keyword = subject
Ciphertext = eakjnqvrqvbrjmouiqeajspifnogonnpn

After that we split the ciphertext into two parts ,the first part = 'eakjnqvrqvbrjmou ' and the second part ='iqeajspifnogonnpn' . the first part , the new subject , replaces the old subject of the keyword statement whereas the second part , the new object , replaces the object of the keyword statement. The predicate is still the same without replacement in order to use it in the deciphering process. Then the new statement ,shown in figure (7) below, will be "eakjnqvrqvbrjmou subject iqeajspifnogonnpn". Figure (8) shows the Beaufort Cipher algorithm. In Deciphering process the receiver knows that the keyword is the predicate of the less important statement, he can merge the subject and object of this statement to make the ciphertext and he uses it with the predicate to find the original plaintext.

Ciphertext = eakjnqvrqvbrjmouiqeajspifnogonnpn
Keyword = subject
Plaintext = our army begin the attack after one hour

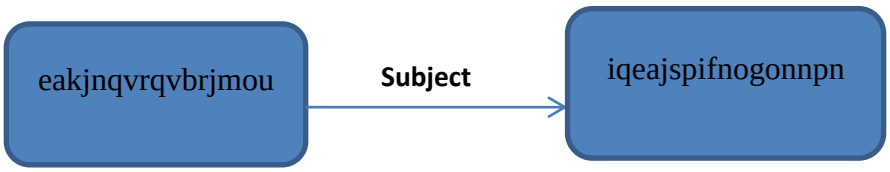

Figure (7) new statement for Alzhemers

```
Input :   Plain text (P) , Keyword (K), Beaufort table (B)
Output : Cipher Text  (C )
Method
Step 1 : S1= length (K)
Step2 : S2 = length (P)
Step3 : Rearrange the keyword (K) according to Plain text length and
        save it to New-Key which its length is the same as plaintext
        length.
 Step4 : For I = 1 to S2 do
Step 5 :  C[i] =  cross-point  between P[i] and k[i]
Step 6 : End for
End Method
```

Figure (8) Beaufort Cipher algorithm

2. Amyloid precursor protein

This RDF file contains (12) Triples : the no. of subject is (1), the number of predicate is (10), the number of object is (12). We take the same plaintext and after analyzing the RDF , we take the less important statement. So the keyword statement is " Amyloid_precursor_protein_secretase type Integral Membrane Proteins". Where the predicate is "type" . Figure (9) show the new statement .

Plaintext = our army begin the attack after one hour
Keyword = type
Ciphertext = feyecmrdpshrarleafpcjyklphbrprbkc

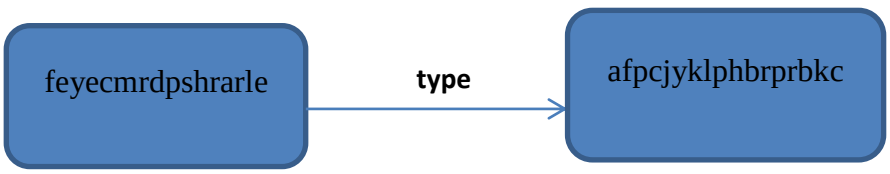

Figure (9) new statement for Amyloid precursor protein

The inverse process can be used in the deciphering process, where we use the same keyword for finding the plain text from the cipher text.

X. Conclusions

We have presented a method for hiding any text information in the Linked Opened Data (RDF representation). We use ciphering techniques to increase the security of our method. The Beaufort cipher is implemented (we can use any

ciphering algorithm). Two types of Linked Opened Data in RDF representation has been used that represents the Alzheimer disease and Amyloid precursor protein. The proposed method can be used in any web site represented by RDF.

References

André Freitas, Edward Curry, João Gabriel Oliveira and Seán O'Riain, 2012, " **Querying Heterogeneous Datasets on the Linked Data Web Challenges, Approaches, and Trends**", IEEE Computer Society.

Arvind Kumar, Km. Pooja,2010, " **Steganography- A Data Hiding Technique** ",*International Journal of Computer Applications (0975 – 8887) Volume 9– No.7.*

Bo Leuf , 2006, **"The semantic web:crafting infrastructure for agency "**, book, John Whiley & Sons,Ltd.

Bret Dunbar, 2002**,"A Detailed look at Steganographic Techniques and their use in an Open-Systems Environment"**, SANS Institute.

Christian Bizer, Jens Lehmann, Georgi Kobilarov, Soren Auer,Christian Becker, Richard Cyganiak, Sebastian Hellmann, 2009, **" DBpedia - A crystallization point for theWeb of Data"**, Web Semantics: Science, Services and Agents on theWorldWideWeb 7 154–165.

Christian Bizer, 2009, **"The Emerging Web of Linked Data "**, IEEE Computer Society.

Christian Bizer, Tom Heath, Tim Berners-Lee,2009, **" Linked Data-The Story So Far"**, International Journal on Semantic Web and Information Systems (IJSWIS), Vol. 5, No. 3, pp. 1-22.

Dominic Hughes, Vitaly Shmatikov, **" Information Hiding, Anonymity and Privacy:A Modular Approach"**, 2004, Journal of computer security, vol.12, issue 1.

Edgard Marx , Percy Salas , Karin Breitman , José Viterbo an Marco Antonio Casanova , 2012, **" RDB2RDF: A relational to RDF plug-in for Eclipse"** , Software- Practice and Experience.

Eric Miller,1998, **" An Introduction to the Resource Description Framework"**, Bulletin of the American Society for Information Science.

Greg kipper, 2004, **" Investigator's Guide to Steganography"** , book, ISBN:0849324335.

Michael Hausenblas, 2009, **"Exploiting Linked Data to Build Web Applications"**, IEEE Computer Society.

Michael Hausenblas, Marcel Karnstedt, 2010, **" Understanding Linked Open Data as a Web-Scale Database "**, IEEE Computer society.

Richard Sheffield, 2009, "The Web Content Strategist's Bible ", book, CLUEfox Publishing.

Sabu M Thampi, 2004, **"Information Hiding Techniques: A Tutorial Review"**, ISTE-STTP on Network Security & Cryptography, LBSCE.