Towards a Language-Independent Universal Digital Library

Sameh Alansary*†
Sameh.alansary@bibalex.org

Magdy Nagi*††
magdy.nagi@bibalex.org

Noha Adly*††
noha.adly@bibalex.org

* Bibliotheca Alexandrina, P.O. Box 138, 21526, El Shatby, Alexandria, Egypt.

† Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University , El Shatby, Alexandria, Egypt.

†† Computer and System Engineering Dept. Faculty of Engineering, Alexandria University, Alexandria Egypt.

Abstract — **Nowadays a dramatic change made by Information Technology (IT) to the concept of libraries emerged. Libraries are gradually going towards digitization which allows for a permanent accessibility of human's work to billions of people all over the world. Nevertheless, digitization alone does not necessarily lead to 'universality'.**

One of the main principles of a universal digital library is its independency of both time and location. This paper adds another dimension to the concept of universality of the digital library, which is 'language-independency'. It therefore, investigates the meaning of the term 'universal', its relation with 'digital' and introduces its realization within the Universal Networking Language (UNL) framework. The paper also describes the role of Ibrahim Shihata Arabic UNL Center (ISAUC) in building language-independent materials in UNL format from Arabic texts. A prototype of a UNL-based Library Information System (UNL LIS) has been implemented as a proof of the concept of 'language independency'.

Index of terms — **Interlingua, UNL, Universal Digital Library, Arabic Natural Language Processing, Language-independency, Library Information System.**

I. INTRODUCTION

In the current digital environment age, IT has been used extensively to record, store, and disseminate information digitally. IT has almost converted the world into a global village. Making the full text of libraries' assets available digitally is a natural step in widening access to many collections.

Libraries are also changing to meet the demand put on them. The new generation whose demand for information is never met is always demanding that traditional libraries should be developed as well-equipped and interconnected digital libraries. Information is for use and for all, accordingly libraries should be for all. Libraries should then become universal for all those who seek information.

The shift to digital libraries continued without a stop mainly because of the so many advantages of a digital library that covered its disadvantages. Some of the digital libraries' advantages are:

- Being independent of time and location: users do need not go to the library physically; people from all over the world can gain access to the same information, also people can gain access to the information at any time, night or day;
- Multiple accesses of same resources;
- Providing access to much richer content in a more structured manner, i.e. we can easily move from the catalog to the particular book then to a particular chapter and so on;
- Information retrieval: providing the possibility to retrieve information from full text;
- Preservation and conservation: an exact copy of the original can be made any number of times without any degradation in quality; and
- Networking: a particular digital library can provide a link to any other resources of other digital libraries very easily; thus a seamlessly integrated resource sharing can be achieved.

Moreover, the issue of 'universality' is vital for a digital library to maintain and preserve its advantages and importance.

There have been several attempts to describe the 'universal' encyclopedia or 'universal' library. In the 17th century, the French librarian Gabriel Naudé has made one of these attempts since the idea of universality continued to be attractive. Naudé considered the universal library to be that library which contains all existing, possible, or useful information or knowledge [5]. This definition of the term 'universal' is familiar to 'library' because long ago such a library was built. The Library of Alexandria, constructed around 300 B.C., was designed to hold all the

1

scrolls circulating in the world. It is regarded as the first library approaching universality in this sense.

In modern times, the idea of a universal library expresses itself in massive digitization projects, such as that of the Open Content Alliance (OCA), Google Print, the Million Book Project [12], Carnegie-Mellon's Universal Library [2] and the European project Bibliotheca Universalis [17]. However, many technical and legal problems remain for the dissemination of all possible knowledge on the Internet. According to old times 'universal' meant 'exhaustiveness' i.e. to have a library that includes collections in all fields of knowledge, while in modern times, 'universal' denotes the antonym of 'local' in addition to 'comprehensive'.

Where the meaning of 'universality' denotes the digitization of books and making them available in a manner that is independent of location and time, the paper in-hand is concerned with adding a very important dimension to 'universality' which is independency of language. This dimension has not been mentioned or stated by anyone as of its advantages.

The paper is organized as follows: Section II discusses how 'language dependency' blocks information dissemination and how Machine Translation (MT) systems have some drawbacks to deal with the situation. Section III presents the issue of 'language independency' as a new dimension to the concept of 'universality'. Section IV presents the UNL system as a universal system for knowledge representation which can act as a framework for the realization of universality for digital libraries . It also presents the contribution of Ibrahim Shihata Arabic UNL Center (ISAUC) in building the infrastructure of Arabic language tools for encoding Arabic texts to universal representation and decoding any universal representation to Arabic. This section ends by presenting a UNL application as a proof of concept, namely, Library Information System that interrogates a language-independent database using Arabic free input in addition to other few natural languages. Finally, section V concludes the paper and discusses future work.

II. LANGUAGE-DEPENDENCY BLOCKS INFORMATION DISSEMINATION

All digital libraries are language-dependent. Being language-dependent, the digital materials will only be read in one's native language. If you are a native speaker of Arabic and you want to read an English e-book, you can only read it in English which will hold a barrier between Arabic and English as a result of language-dependency.

To break language barriers, translation systems have been introduced. They are designed either for one particular pair of languages (bilingual systems) or for more than two languages (multilingual systems), either in one direction (uni-directional systems) or in both directions (bi-directional systems) [10], [11]. In overall system design of MT systems, three basic types can be distinguished. The first is generally referred to as the *'direct translation'* approach: the MT system is designed in all details specifically for one particular pair of languages in one direction [6], [7]. In this case, source texts are analyzed no more than necessary for generating texts in the other language ([4], [13], and [11]). The second type is the *'transfer approach'*. In this type three stages can be underlined. The first stage converts texts into intermediate representations, usually syntactic representations for both source and target texts, in which ambiguities have been resolved irrespective of any other language. In the second stage these are converted into equivalent representations of the target language; and in the third stage, the final target texts are generated. Analysis and generation programs are specific for particular languages and independent of each other. Differences between languages, in vocabulary and structure, are handled in the intermediary transfer program [7].

The third basic type is the *'interlingual'* approach, which assumes the possibility of converting texts to and from 'meaning' representations common to more than one language. Translation is thus obtained in two stages: from the source language to the interlingua, and from the interlingua into the target language. Interlingual machine translation is one instance of rule-based machine translation approaches. According to this approach, the source language, i.e. the text to be translated is transformed into an interlingual, i.e. language independent representation. The target language is then generated out of the interlingua [6],[7],[8].

The interlingual approach to machine translation has advantages and disadvantages. The advantage is that no transfer component has to be created for each language pair. The obvious disadvantage is that it is difficult and maybe even impossible for a wider domain. The ideal context for interlingual machine translation is thus multilingual machine translation in a very specific domain. However, large-scale interlingual MT systems have been constructed and been very effective, the best known being the Fujitsu system in Japan (http://www.fujitsu.com/global/services/translation). Early interlingual MT systems were built at Stanford in the 1970s by Roger Schank and Yorick Wilks [8]. The former became the basis of a commercial system or the transfer of funds, and the latter is preserved in the Unites States' Computer Museum at Boston as the first interlingual machine translation system. Fig. 1 shows a comparison

between the three approaches as far as intermediary representation is concerned: interlingual machine translation at the peak, followed by transfer-based, then direct translation.
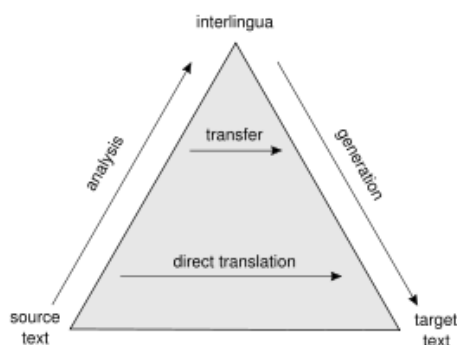

Fig. 1. A pyramid showing comparative depths of intermediary representation.

Several multilingual translation systems such as Google (http://www.google.ch/language_tools) and Altavista (http://babelfish.altavista.com) have been created to solve language problems on the Internet. On one hand, these kinds of systems have significant weakness and in most of the cases the quality of results is often inadequate and only work for a limited number of language combinations (cf. [1] for more information about drawbacks of other systems). On the other hand, even if we suppose that a multilingual system that translates from and to all languages exists, this will imply huge language resources such as grammars, lexicons and translation dictionaries which might hold an overload on the network. For example, if a multilingual system translates from and to only 10 languages, 10 grammars, 10 lexicons, 90 translation dictionaries and 90 sets of translation rules will be needed, plus the need for semantic processing in each language. It is, thus, recommended to shift towards language-independent machine translation systems, or interlingual machine translation systems.

It is believed that if digital libraries are language-dependent, this will block information dissemination which will in turn endanger the concept of 'universality' i.e. language-dependency dimension adds a degree of 'locality' rather than 'universality' to digital libraries in particular and to digitization in general.

III. 'UNIVERSALITY' IS THE INDEPENDENCY OF TIME, PLACE AND LANGUAGE

As has been mentioned in the last section, language-dependency represents an obstacle in disseminating information which makes the universality of digital libraries to some extent limited. Therefore, if the dimension of language-independency has been added to a Universal Digital Library, more universality will be achieved that goes beyond language barriers. One question might appear here: what is the optimum sense of "language-independency", which adds universality to digitization?

The optimum sense of 'language-independency' is the independency of whatever natural language the electronic text is written in. In a language-independent digital library, search will not depend on the language in which the search word is written in. The search engine, therefore, will rather bring results from all different natural languages.

One of the recent major contributions of the Universal Digital Library to the Indian language information technologies for the representation of knowledge is the Multilingual Book Reader Interface [3]. This system makes the digital library of India available to anyone at anytime and from anywhere. The main aim of this work was to add the dimension of any-language to the system. The book reader allows any presented electronic text in any Indian language to be transliterated into any one of the many Indian languages. The provided features by the interface are advantageous not only to the readers who can understand but not read their own language, but also to those who desire to obtain at least a simple translation of the book to their desired language.

The work in [3] represents a pioneering work to language-independency, however, it allows only for the translation or transliteration between Indian languages which added a degree of nationality to the interface. Accordingly, it did not cover fully the dimension of universality the current paper seeks to present.

This leads us to highly recommend shifting from language-dependent digital libraries to language-independent ones based on the interlingual concept. Language-independent digital libraries could then be housed by the Internet and accessed in one's mother tongue which helps to improve the global society in several ways beyond measurement.

IV. TOWARDS A UNIVERSAL SYSTEM FOR KNOWLEDGE REPRESENTATION

In order to represent natural language materials in a language independent format, a system is required to represent knowledge of texts in a universally digital format given the conditions that: a) the content (meaning) of the original materials must not be lost; b) this universal format should be understandable by various platforms over the network (internet). One of the recent major contributions of the Universal Networking Digital Language Foundation (UNDL) is the UNL system. The

UNL system, which has been developed under the UNL Programme started in 1996, as an initiative of the UNU/IAS, provides an interlingual universal system for knowledge representation aiming at breaking the barriers between world's languages. The knowledge expressed by UNL is universal and can be shared by all nations. With UNL, computers can reason using knowledge which will give possibility to a more intelligent processing such as semantic computing. Semantic computing, in turn, will lead to building an intelligent search engines allowing users to search books' contents which would be other wise unfeasible.

A. The UNL System

The Universal Networking Language (UNL) is an artificial language for computers to express information and knowledge that can be expressed in natural languages [16]. The UNL has been developed to express any kind of knowledge that can be represented in any natural language and to make it possible that such knowledge can be fully accessible in any natural language. As a language for representing information and knowledge described in natural languages, UNL has all the components corresponding to that of a natural language. It is composed of words expressing concepts called 'Universal Words' (UWs) that are inter-linked with other UWs to form the UNL expressions of sentences, the UNL graph [14] . Fig. 2 shows a UNL graph of the sentence: 'I hear a dog barking outside'. These links, called 'relations', specify roles of each word in a sentence.
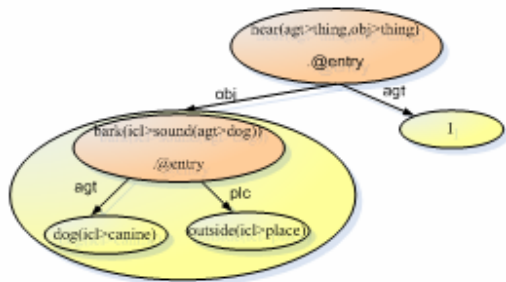
Fig. 2. A UNL graph

The subjective meanings intended by the author are expressed through 'attributes' which can convey information such as the time with respect to the speaker, the speaker's view of aspect, reference, attitude, emphasis and focus. Its UW (Universal Word) System has been developed allowing people to introduce and define any concept (UW) no matter how particular or specific to a

language it is. All UWs defined in the UW System can be understood by computers by describing all possible relations that every UW has with other UWs in UNL Knowledge Base (UNLKB). Such UNLKB not only provides linguistic knowledge in the form that computer can understand but also it represents a semantic network defining every possible binary relation between concepts (i.e. between UWs), as shown in Fig. 3. Its entries follow the format 'relation(UW1, UW2)=c', where 'relation' stands for any UNL Relation, 'UW1' and 'UW2' are two different UWs, and 'c' is the degree of certainty, ranging from 0 (impossible) to 255 (absolutely certain).
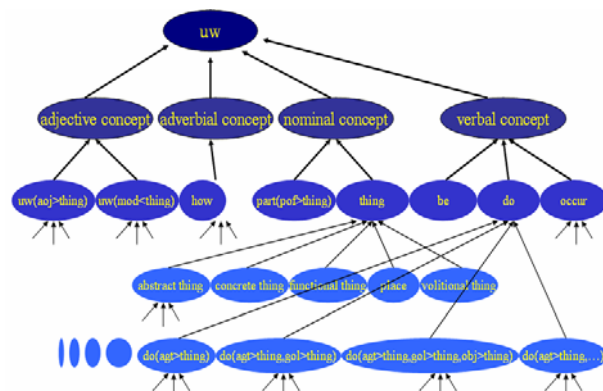
Fig. 3: The UNL Knowledge Base system

The UNL System (Fig. 4) is designed so that it can be used over the internet. It consists of multiple language servers – one language server for each natural language – as well as UNL Editors and UNL Viewers. Pages written in any language can be viewed in the user's natural language by a simple process of "EnConversion" and "DeConversion". The language server for each language is responsible for converting a natural language into UNL (EnConversion) and UNL into a natural language (DeConversion). UNL Editors and UNL Viewers assist the user to create and view UNL documents, respectively. For instance, when a page is being developed in Arabic, the UNL Editor sends a request to the Arabic language server to "enconvert" the page to UNL. To view this page in Spanish, for example, the UNL Viewer sends a request to the Japanese language server to "deconvert" the UNL contents into Spanish.
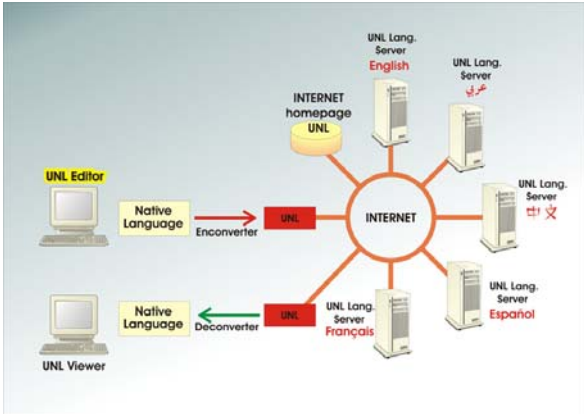
4

Fig. 4. The UNL system

A language server consists of a collection of software components that listen for conversion requests, perform the conversion and send the converted text back to the user. The software components include an Enconverter and a Deconverter. Both the Enconverter and the Deconverter make use of two files in their conversion processes, a dictionary file and a rule file. The dictionary file contains a list of natural language words and their corresponding Universal Words of the UNL. The rule file contains a list of grammatical rules that govern either how a natural language sentence is converted into UNL (EnConversion rules) or how a UNL sentence is converted into a natural language (DeConversion rules).

B. UNDL and ISAUC

The Universal Networking Digital Language Foundation (UNDL) has 15 partners worldwide. Each partner is responsible for building and maintaining a language server for their particular language. This entails compiling a Universal Word dictionary, building the DeConversion as well as EnConversion rules for that language, in addition to updating the UNL KB with their newly created concepts.

Ibrahim Shihata Arabic-UNL Center (ISAUC), established at the Bibliotheca Alexandrina, is responsible for designing, implementing and maintaining the various components of the Arabic Language Server. Although we are still at an early stage of building the infrastructure of the Arabic Language Server, progress has been achieved in the following areas:

1) The Arabic Dictionary: It is the repository of information for all UNL Arabic grammars. Therefore, much attention has been given to the dictionary in order to make it ready in the required format that supports morphological, syntactic and semantic analysis and generation needed for both Arabic EnConversion and DeConversion rules. The design of the dictionary includes a stem based Head Words (HW) capable of representing all paradigms of the Arabic lexeme, the Universal Concept that this HW represents and other information that defines HWs on various linguistic levels representing their morphological attributes, the syntactic slot they can fulfill and idiosyncratic semantic properties that govern UNL relations.

2) Arabic DeConversion rules: These rules are responsible for generating Arabic sentences out of UNL networks, the language independent representation of texts. This module should therefore be able to select the Arabic words that represent universal concepts, arrange the concepts of the UNL network in a syntactically well-formed sentence, and prevent any deviation in meaning of the generated sentence from the original sentence, the sentence from which the UNL network has been EnConverted.

3) Arabic Enconversion: This module is responsible for encoding Arabic sentences to UNL networks. It is capable of performing morphological analysis to extract concepts the Arabic words refer to, assign exact semantic relation between concepts as being expressed in the context of the Arabic sentence. This module is quite challenging as it is supposed to be working with Arabic sentences in unrestricted domains.

4) A corpus for Modern Standard Arabic (MSA): It is a representative sample that reflects the empirical usage of MSA. Information extracted from the Arabic corpus will play a principal role in enhancing and updating both EnConversion and DeConverstion rules. The MSA corpus is built using an application developed by the ISAUC.
Having tested the Arabic language tools (the dictionary and the conversion grammars) against adequacy, the Arabic language server will be set up to act as an active center for representing Arabic digital assets in a universal format for other language center to deconvert. Also, the Arabic language server will be capable of deconverting the universal materials produced by other language centers to Arabic.

Some tools and applications utilizing the UNL framework have also been developed, such as the UNL Library Information System (UNL LIS) and the UNL Integrated Development Environment (UNL IDE). The UNL LIS is an application that allows the retrieval of information about books available in a data store. It will be explained in detail in the next section. The UNL IDE is a tool that enables users and developers to view the UNL semantic network, search the UNL documents, write rules, check their syntax, and debug and watch the Deconversion

5

and Enconversion outputs for the given rules and dictionary.

In its current state the Arabic language server can EnConvert Arabic materials to universal representation in UNL format as far as the efficiency of the dictionary and the grammars is concerned. For example, consider the Arabic sentence in (1), taken from the biography of the ex-president Gamal Abdel Nasser that exists within 'Nasser digital library' developed by Bibliotheca of Alexandria (http://nasser.bibalex.org/pictures01-%20sira.htm):

وكان جمال عبد الناصر الابن الأكبر لعبد الناصر حسين الذي (1)
ولد في عام 1888 في قرية بني مر في صعيد مصر في أسرة
من الفلاحين، ولكنه حصل على قدر من التعليم سمح له بأن
يلتحق بوظيفة في مصلحة البريد بالإسكندرية، وكان مرتبه
يكفي بصعوبة لسداد ضرورات الحياة.

'Gamal Abdel Nasser was the elder son of Abdel Nasser Hussain who is born in 1888 in Bani Morr Village in Upper Egypt in a family of farmers, but he has got a degree of education that allowed him to join a job in the postal service in Alexandria, and his salary was hardly sufficient to satisfy life demands'

This sentence passes through a morphological analysis stage to extract concepts (UWs) represented by the Arabic words (Fig. 5) then these concepts are linked together by semantic relations to build the UNL hyper semantic network as shown in Fig. 6 which can be represented graphically as in Fig. 7

Gamal Abdel Nasser(iof>person):00
son(icl>person):0I.@def
most(icl>how):15
old(aoj>thing):1J
Abd El-Naser Hossain(iof>person):23
born(obj>thing):31.@past
year(icl>period):3M, 1888:41
village(icl>region):4D
Bani Morr(iof>village):4S
upper Egypt(iof>place):58
family(icl>group):5Q
farmer(icl>person):65.@pl.@def
get(agt>thing,obj>thing):6S.@past.@contrast
degree(icl>abstract thing):7N
education(icl>activity):82.@def
allow(agt>thing,gol>thing,obj>thing):8M.@past
his(pos>he):97
join(agt>person,obj>thing):9I.@present
job(icl>work):A7
postal service(icl>service):AN
Alexandria(iof>city):BB
salary(icl>money):BV
his(pos>he):CB
suffice(aoj>thing,obj>thing):CM.@present
hardly:DA
satisfy(agt>thing,obj>thing):DQ
demand(icl>wants):E6.@pl.@def
life(icl>activity):EV.@def

Fig. 5 Concepts extracted from the Arabic sentence in (1).

```
{unl}
aoj(son(icl>person):0I.@def.@entry,        Gamal Abdel Nasser(iof>person):00)
mod(son(icl>person):0I.@def.@entry,        Abd El-Naser Hosen(iof>person):23.@topic)
aoj(old(aoj>thing):1J,      son(icl>person):0I.@def)
man(old(aoj>thing):1J,      most(icl>how):15)
obj(born(obj>thing):31.@past,      Abd El-Naser Hossain(iof>person):23.@topic)
and(get(agt>thing,obj>thing):6S.@past.@contrast,      born(obj>thing):31.@past)
scn(born(obj>thing):31.@past,      family(icl>group):5Q)
plc(born(obj>thing):31.@past,      village(icl>region):4D)
tim(born(obj>thing):31.@past,      year(icl>period):3M)
mod(year(icl>period):3M, 1888:41)
plc(village(icl>region):4D,        upper Egypt(iof>place):58)
mod(village(icl>region):4D,        Bani Morr(iof>village):4S)
mod(family(icl>group):5Q,          farmer(icl>person):65.@pl.@def)
obj(get(agt>thing,obj>thing):6S.@past.@contrast,      degree(icl>abstract thing):7N)
agt(allow(agt>thing,gol>thing,obj>thing):8M.@past, degree(icl>abstract thing):7N)
mod(degree(icl>abstract thing):7N, education(icl>activity):82.@def)
gol(allow(agt>thing,gol>thing,obj>thing):8M.@past, join(agt>person,obj>thing):9I.@present)
obj(allow(agt>thing,gol>thing,obj>thing):8M.@past, his(pos>he):97)
and(suffice(aoj>thing,obj>thing):CM.@present,      join(agt>person,obj>thing):9I.@present)
obj(join(agt>person,obj>thing):9I.@present, job(icl>work):A7)
plc(job(icl>work):A7,      postal service{icl>service ):AN)
plc(postal service{icl>service ):AN,          Alexandria(iof>city):BB)
aoj(suffice(aoj>thing,obj>thing):CM.@present,      salary(icl>money):BV)
mod(salary(icl>money):BV,      his(pos>he):CB)
obj(suffice(aoj>thing,obj>thing):CM.@present,      satisfy(agt>thing,obj>thing):DQ)
man(suffice(aoj>thing,obj>thing):CM.@present,      hardly:DA)
obj(satisfy(agt>thing,obj>thing):DQ,        demand(icl>wants):E6.@pl.@def)
mod(demand(icl>wants):E6.@pl.@def,      life(icl>activity):EV.@def)
{/unl}
```

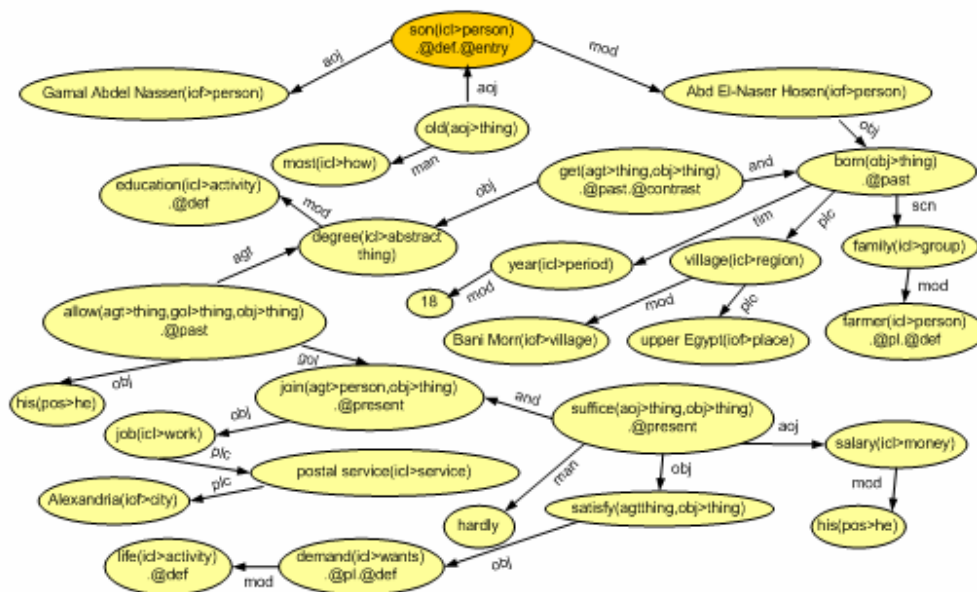Fig. 6. The UNL network generated from the sentence in (1)

Fig. 7. Graphical representation of the UNL network in Fig. 6.

7

C. The UNL Library Information System (UNL LIS)

As it will take a few years to set up the Arabic language server which will play the main role in making the concept of language independency feasible through the network (Internet) which will in turn make the concept of 'language-independent Universal Digital Library' a reality, a UNL-based Library Information System has been implemented as a proof of concept. The UNL LIS provides access for its users to books of different languages in their native language. The users have access to both the metadata and the contents of the books. They can search the book catalog by entering a query in their natural language inquiring about metadata such as the title or the author of a certain book, or even books by a certain author. The query is EnConverted to UNL and the search is performed on the ensemble of books residing in the system. The search results contain the metadata of the books matching the query. These search results are then DeConverted by calling the corresponding language server and displayed to the user in his native language. Furthermore, the contents of the books are also converted to UNL and added to the UNL Encyclopedia, which contains the set of all UNL documents. The contents of these books can also be deconverted and displayed in any natural language. Fig. 8 illustrates the system architecture of UNL LIS.
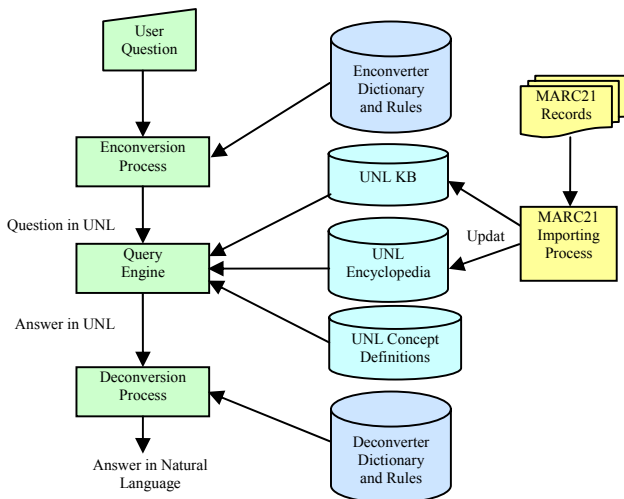

Fig.8. The System Architecture of UNL LIS

MARC21 records are the standard format used by the Bibliotheca Alexandrina to catalog books and store their metadata. An application is used to import the MARC21 records into the UNL Knowledge Base (KB) and the UNL Encyclopedia.

Since the UNL formalism is associated with a very abstract level of language representation, i.e. concepts, a detailed computational linguistic techniques have been used to deal with different possible interrogative Arabic structures. Therefore, it was possible to reach a degree of reliability of what so called "question understanding" rather than "pattern matching".

The system starts by receiving a query in natural language which is enconverted to UNL. The Query Engine receives the question in UNL, searches the system and generates an answer in UNL which is then fed to the DeConverter. To answer a question, the Query Engine needs three sources of information, the UNL KB (Knowledge Base), the UNL concept definitions and the UNL Encyclopedia. The UNL KB provides semantic information for words that help in understanding natural language sentences. For example, it might contain information such as an "author" is a "person". The UNL concept definitions define UNL concepts in terms of other concepts, for example "a person who writes a book" is the definition of an "author". The third source of information, the UNL Encyclopedia, makes available all kinds of knowledge linked to every concept.

The Query Engine is implemented in Prolog. Using Prolog the process of query answering is actually a search of possible interpretations of the question. As mentioned, the input to the query engine is a set of UNL relations. Let us consider the Arabic question "من مؤلف السكرية؟" 'Who is the author of Al Sukaria?' represented by the following UNL that is obtained by means of the Arabic Enconversion rules:

```
 mod(author(icl>person):03.@entry.@interrogative,
Al Sukaria( iof>book,mod>Naguib Mahfouz):08)
        iof(who(icl>person):00.@qfocus,
author(icl>person):03.@entry.@interrogative)
```

To match this query, the engine replaces the question focus node (such as who, where, when, which) with an appropriate variable. In this case, the node "who(icl>person)" will be replaced by the variable "Person". The resulting graph is then matched against all the documents of the UNL Encyclopedia one relation at a time, Fig. 9.
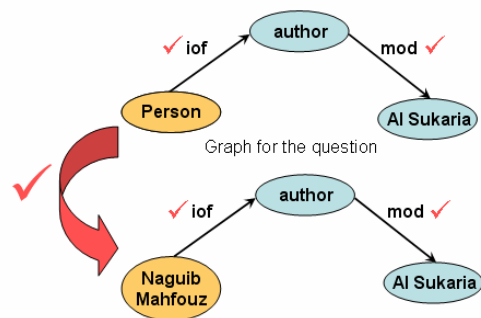

Fig.9. An example of a direct match

If, however, the question was rephrased as " من الذي ألف السكرية؟" 'Who authored Al Sukaria?' and entered to the system, the Query Engine will not find a match instantly.

The UNL LIS application is a web application. Snapshots of the results of a user query in Arabic is shown in Fig. 10. As shown, the question " من الروائي الذي كتب السكرية؟" 'who is the novelist who wrote Al Sukaria?' is fed to the UNL LIS. The system displays the answer " نجيب محفوظ" 'Naguib Mahfouz' which is also a link to further information about the results. The biographic data of Naguib Mahfouz can be displayed, as in Fig. 11, in the selected language, Arabic in the current case.

This biography has been generated from a language independent representation.

It will thus transform the question back to "Who is the author of Al Sukaria?" by applying the inference rules stored in the UNL Concept Definitions.

At present, answers can be deconverted into six more languages in addition to Arabic. The languages are: Armenian, Chinese, English, Indonesian, Japanese, and Portuguese. The system handles questions in a free language input in both Arabic and English (600 questions have been tested from native speakers of Arabic and English). A subset of these questions has also been implemented by the Indonesian and Japanese language centers.

Fig. 10. Query results

Fig. 11. An example of generation of the biography of Nagib Mahfouz from a language independent representation.

9

V. CONCLUSION AND FUTURE WORK

It is very important that Digital libraries reach out to users from everywhere, serving information needs at all levels. To do this, they should attain the concept of universality. The present paper argued that it is extrapolated that three dimensions are necessary for the production of competent digital library systems capable of disseminating knowledge. These dimensions are the independency of time, location and language.

The UNL system is a promising formalism for representing knowledge that is expressed in natural language. It allows for expressing any specific concept of a particular language precisely without ambiguities and understandable for computers and humans of other languages. This mechanism represents a strong basis upon which all languages can have interfaces with UNL and makes it possible that anyone can participate in the development of a language module or UNL-based application which can serve the purpose of universal digital libraries.

We presented the contribution of the ISAUC in designing and implementing the Arabic language components of the UNL system. Also, we have presented an application based on UNL to prove the feasibility of the concept of language independency. The UNL LIS application allows users to both interrogate a UNL-based library catalogs in their native languages and receive answers in natural language. The application currently supports questions in two languages (Arabic and English) and answers can be displayed in 7 languages.

REFERENCES

[1] Alansary, S. (2004), "Arabic-English Machine Translation Systems: Discrepancies and Implications", *JEP/TALN International Conference, Special session on Arabic text and speech language processing*. Fez, Morocco, 19-22 April 2004.

[2] Alper, J. (1998), "Digital Libraries: Assembling the World's Biggest Library on your Disktop", *Science*, Volume 281, Number 5384, pp. 1784 – 1786.

[3] Balakrishnan, N. (2006), "Universal Digital Library Connecting Users to Digital Contents", *13th Biennial Conference and Exhibition, VALA (2006),* Melbourne, Australia.

[4] Bennett, P. A., Johnson, J. Mc, Naught, J. M., Pugh, JC. Sager and H.L. Somers (1986), "*Multilingual aspects of information technology",* Gower Publishing Limited, Aldershot.

[5] Hedstrom, M. and King, J. (2004), " *On the LAM: Libraries, Archives, and Museums Collection in the creation and maintenance of knowledge communities". Paper published online by Organization for Economic Co-operation and Development(OECD:*
http://www.oecd.org/dataoecd/59/63/32126054.pdf

[6] Hutchins, W. J. (1986) *Machine Translation: Past, Present,Future*. Ellis Horwood. Chichester.

[7] Hutchins, W. John (2003). Machine Translation: General Introduction. In Ruslan Mitkov (2004), *"The Oxford Handbook of Computational Linguistics"*. Oxford University Press.

[8] Hutchins, W. John and Harold, L. Somers (1992), *"An Introduction to Machine Translation"*, London: Academic Press.

[9] Johnson, R. L. (1984), *"Contemporary perspectives in machine translation",* Contrasts (Revue de l'Association pour le Developpement des Etude Contrastives) hors serie A4 "Traduction Automatique- Aspects Europeens", pp.141-155, janvier.

[10] King, M. (1982), "Eurotra: An Attempt to Achieve Multilingual MT". In Lawson (1982) (ed.): *Practical experience of machine translation,* pp.139-148. North-Holland publishing company, Amsterdam.

[11] Lehrberger, J. and L. Bourbeau (1988), *"Machine translation: linguistic characteristics of MT systems and general methodology of evaluation",* John Benjamins, Amsterdam.

[12] Reddy, R. and Carbonell j. (2001), *"The Million Book Digital Library Project"*, Technical report, Carnegie Mellon University. Http://www.rr.cs.cmu.edu/mbdl.htm .

[13] Tucker, A.B. (1987), "Current strategies in machine translation research and development", In Nirenburg (1987) (ed.), *Machine Translation: Theoretical and Methodological Issues*, pp. 22-41, Cambridge University Press.

[14] Uchida H. (2002), *"How to Build Universal Knowledge",* International Conference on Universal Knowledge and Language, ICUKL*, Goa, India.

[15] Uchida H., Zhu M. (2002), *"Universal Word and UNL Knowledge Base"*, *International Conference on Universal Knowledge and Language*, ICUKL, Goa, India.

[16] Uchida H., Zhu M. (2005), "UNL2005 for Providing Knowledge Infrastructure", *Workshop on the Semantic Computing Initiative (SeC2005)*, Chiba, Japan.

[17] Zillhardt, S. (1998), "Bibliotheca universalis : a G-7 global information society pilot project" *64th IFLA General Conference. Amsterdam*