Machine Translation Using the Universal Networking Language (UNL)

Sameh Alansary[1,2] Magdy Nagi[1,3] Noha Adly[1,3]
Sameh.alansary@bibalex.org magdy.nagi@bibalex.org noha.adly@bibalex.org

[1]Bibliotheca Alexandrina, P.O. Box 138, 21526, El Shatby, Alexandria, Egypt.
[2]Department of Phonetics and Linguistics, Faculty of Arts, Alexandria University, El Shatby, Alexandria, Egypt.
[3]Computer and System Engineering Dept. Faculty of Engineering, Alexandria University, Alexandria,Egypt.

Extended Abstract

There is a growing insight that high-quality NLP applications for information access are in need of deeper semantic analysis. Many machine translation systems have been developed in the past years. However, these systems faced many problems, some of which are: lexical ambiguity, syntactic ambiguity, referential ambiguity and differences in conceptual specificity.

Machine translation has been brought to a large public by tools available on the Internet, such as Google, AltaVista, and by low-cost programs such as Babylon. These tools produce a "gisting translation" — a rough translation that "gives the gist" of the source text, but is not otherwise usable. MT systems can produce translations more quickly and often more cheaply than human translators; however, in the majority of cases, the quality of MT is inferior to the quality of human translation (HT). An automatic semantic translation could be closer in quality somehow to the quality of an HT. Most machine Translation (MT) systems represent a machine translation as an automatic process that translates from one human language to another language by using context information. However, if the translation stems from a semantic representation the situation will be different.

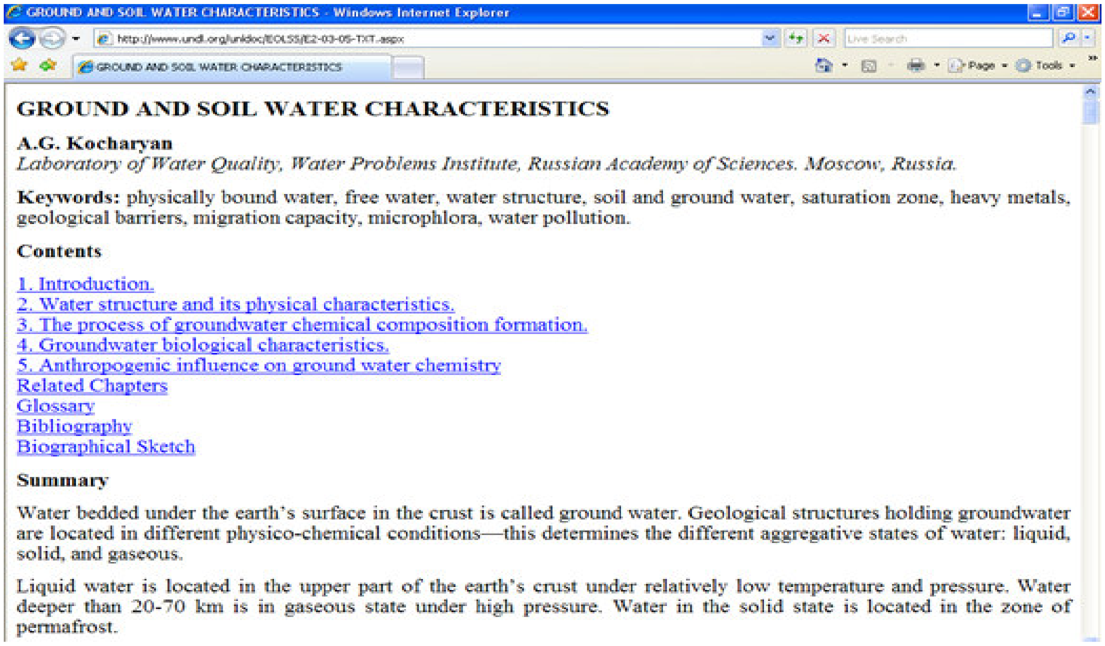
Semantic translation is carried out with reference to grammatical deep structure and it aims at establishing semantic equivalence. If a translation follows the path through semantic representations, it can demonstrate how sentences in the source language and target language relate to a common deep structure. In this method, the machine translation tries to reproduce the precise contextual meaning of the author within the bare syntactic and semantic constraints of the target language. Semantic method emphasizes the content of the message rather than the effect. As it does not want to miss any semantic nuance.

The adopted approach in the translation in this abstract follows a different way, it translates from an Interlingua to different human languages, this Interlingua representation is semantically based. This approach is called “Universal Networking Language (UNL)”. UNL has been introduced by the United Nations University, Tokyo, to facilitate the transfer and exchange of information over the internet. It is an interlingua-based framework that facilitates semantic processing of natural languages by a computer called Universal Networking Language (UNL). An artificial language describes the meaning of sentences in terms of the schema of semantic nets. This framework focuses on representing all sentences that have the same meaning in all natural languages using a single semantic graph. Once this graph is built, it is possible to decode it to any other language.UNL is an electronic language that enables to rewrite articles in any languages on Internet into UNL format in order to translate them into any other languages. It is an interlingua-based framework aimed to facilitate semantic processing of natural language by a computer. Its main applications cover not only in machine translation and other natural language processing tasks, but also in a wide variety of applications ranging from e learning platforms to management of multilingual document bases.

The UNL is language-independent; it provides the possibility to work at the semantic level, it follows the schema of semantic nets-like structure in which nodes are word concepts and arcs are semantic relations between these concepts. It is an Interlingua for machine translation. In this scheme, a source language sentence is converted to the UNL form using a tool called the EnConverter. Enconverter is a language independent parser that provides synchronously a framework for morphological, syntactic and semantic analysis. Subsequently, the UNL representation is converted to the target language sentence by a tool called the DeConverter. The DeConverter is a language independent generator that provides a framework for syntactic and morphological generation as well as co occurrence- based word selection for natural collocation. It can deconvert UNL expressions into a variety of native languages, using a number of linguistic data such as Word Dictionary, Grammatical Rules and Co-occurrence Dictionary of each language.

The purpose of introducing the Universal Networking Language (UNL) in communication networks is to achieve accurate exchange of information among different languages and representing a solution to overcome the barriers of linguistic differences.

One of the challenging missions that the UNL system has faces is to translate the Encyclopedia of Life support system(EOLSS) which is the largest on-line Encyclopaedia; it includes more than 120,000 web pages and it increases constantly. The Arabic UNL language center contributed in translating 25 documents of the Encyclopedia of Life Support Systems (EOLSS) to Arabic. The translation has been achieved by building The EOLSS dictionary, a grammar to deconvert the UNL representation to Arabic. Figure 1 shows a screenshot of an English document from the Encyclopedia of Life Support Systems (EOLSS) while figure 2 shows the Arabic output.


GROUND AND SOIL WATER CHARACTERISTICS - Windows Internet Explorer
http://www.undl.org/unldoc/EOLSS/E2-03-05-TXT.aspx
Live Search
GROUND AND SOIL WATER CHARACTERISTICS
Page
Tools

GROUND AND SOIL WATER CHARACTERISTICS

A.G. Kocharyan
Laboratory of Water Quality, Water Problems Institute, Russian Academy of Sciences. Moscow, Russia.

Keywords: physically bound water, free water, water structure, soil and ground water, saturation zone, heavy metals, geological barriers, migration capacity, microphlora, water pollution.

Contents

1. Introduction.
2. Water structure and its physical characteristics.
3. The process of groundwater chemical composition formation.
4. Groundwater biological characteristics.
5. Anthropogenic influence on ground water chemistry
Related Chapters
Glossary
Bibliography
Biographical Sketch

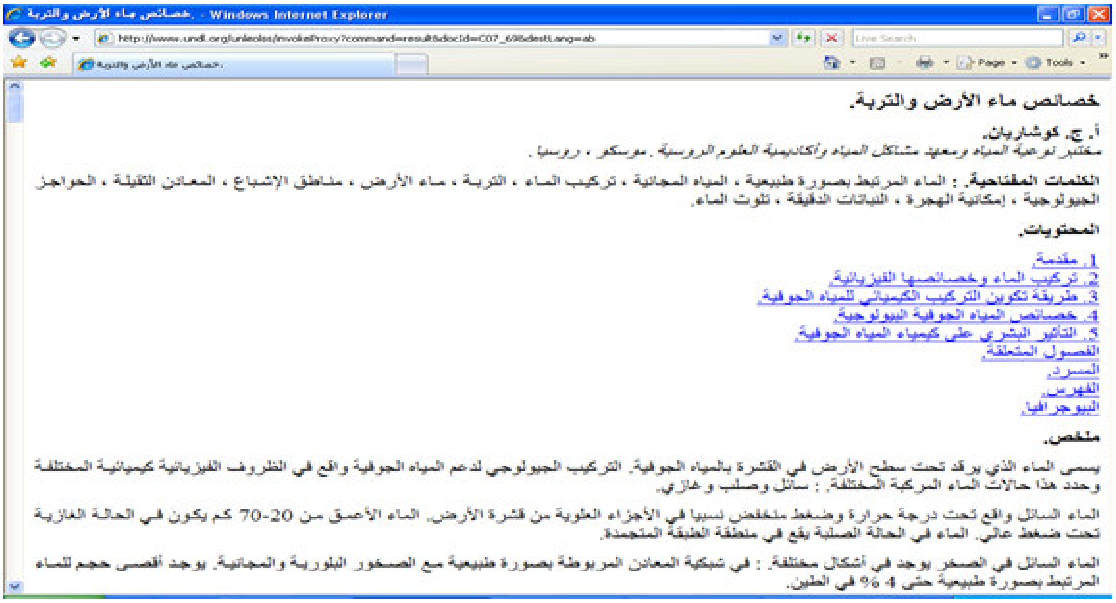
Summary

Water bedded under the earth's surface in the crust is called ground water. Geological structures holding groundwater are located in different physico-chemical conditions—this determines the different aggregative states of water: liquid, solid, and gaseous.

Liquid water is located in the upper part of the earth's crust under relatively low temperature and pressure. Water deeper than 20-70 km is in gaseous state under high pressure. Water in the solid state is located in the zone of permafrost.

Figure 1: A snap shot of an English document

Figure 2 presents a screenshot of Arabic natural language output.


خصائص ماء الأرض والتربة. - Windows Internet Explorer
http://www.undl.org/unleolss/invokeProxy?command=result&docId=C07_69&destLang=ab
Live Search
خصائص ماء الأرض والتربة.
Page
Tools

خصائص ماء الأرض والتربة.

أ. ج. كوشاريان.
مختبر لنوعية المياه ومعهد مشاكل المياه وأكاديمية العلوم الروسية . موسكو ، روسيا .

الكلمات المفتاحية. : الماء المرتبط بصورة طبيعية ، المياه المجانية ، تركيب الماء ، التربة ، ماء الأرض ، مناطق الإشباع ، المعادن الثقيلة ، الحواجز الجيولوجية ، إمكانية الهجرة ، النباتات الدقيقة ، تلوث الماء.

المحتويات.

1. مقدمة
2. تركيب الماء وخصائصها الفيزيائية
3. طريقة تكوين التركيب الكيميائي للمياه الجوفية
4. خصائص المياه الجوفية البيولوجية
5. التأثير البشري على كيمياء المياه الجوفية
الفصول المتعلقة
المسرد
الفهرس
البيوجرافيا

ملخص.

يسمى الماء الذي يرقد تحت سطح الأرض في القشرة بالمياه الجوفية. التركيب الجيولوجي لدعم المياه الجوفية واقع في الظروف الفيزيائية كيميائية المختلفة وحدد هذا حالات الماء المركبة المختلفة. : سائل وصلب و غازي.

الماء السائل واقع تحت درجة حرارة وضغط منخفض نسبيا في الأجزاء العلوية من قشرة الأرض. الماء الأعمق من 20-70 كم يكون في الحالة الغازية تحت ضغط عالي. الماء في الحالة الصلبة يقع في منطقة الطبقة المتجمدة.

الماء السائل في الصخر يوجد في أشكال مختلفة. : في شبكية المعادن المربوطة بصورة طبيعية مع الصخور البلورية والمجانية. يوجد أقصى حجم للماء المرتبط بصورة طبيعية حتى 4 % في الطين.

Figure 2: A snap shot of an Arabic translated document.

The generated Arabic sentences obtained by translating a semantic representation (a data model in the form of universal networking language which consists of words represented by concepts, or universal words (UWs), and the relation between these concepts these are semantic relations in order to convey the correct meaning) has proved highly accurate, and also takes advantage of semantics that associate meaning with individual data elements in the dictionary.

The technical steps followed in evaluating machine translation output included: morphology, i.e. word structure; syntax i.e. the well/ill formedness of the generated sentences, word order, case marking, preposition and particles, and order or modifiers; and semantically, i.e. whether or not the Arabic output still conveys the meaning expressed by the source language. The result of the evaluation is Initial highlighted morphological accuracy of 90%, syntactic accuracy of 75% and semantic accuracy of 85%.

In addition, the semantic representation (UNL expression) form which the Arabic translation in fig. 1 has been generated has a deeper importance in other NLP applications beyond machine translation such as information extraction because in natural language processing the goal of any information extraction system is to automatically extract structured information, i.e. categorized and contextually and *semantically well-defined* data from a certain domain. Further, the use of the universal networking language can contribute to other challenging application as information retrieval, text summarization and Natural language understanding.