TECHNICAL REPORT

ISO/IEC TR 29127

First edition
2011-07-01

Information technology — System Process and Architecture for Multilingual Semantic Reverse Query Expansion

Technologies de l'information — Processus système et architecture pour l'extension multilinguale des requêtes sémantiques inverses

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/IEC TR 29127:2011
https://standards.iteh.ai/catalog/standards/sist/30259af6-29c6-42e0-9cf7-
32184c48cc6b/iso-iec-tr-29127-2011

COPYRIGHT PROTECTED DOCUMENT

Contents

Page

iTeh STANDARD PREVIEW
(standards.itch.ai)

ISO/IEC TR 29127:2011
https://standards.itch.ai/catalog/standards/sist/38259af6-29c6-42e0-9cf7-
32184c48cc6b/iso-iec-tr-29127-2011

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide to publish a Technical Report. A Technical Report is entirely informative in nature and shall be subject to review every five years in the same manner as an International Standard.

ISO/IEC TR 29127 was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 36, *Information technology for learning, education and training*.

Introduction

Learning, Education and Training (LET) in the context of multilingual cultures on a local and global scale can be problematic, especially when learners are proficient in only one language. One of the multilingual problems in a LET environment is how to query LET materials when the requestor cannot understand or is not proficient in the language of the material available.

For example, how does a person who is proficient in French search for, find, and readily understand digital LET materials in Arabic, if the person is not proficient in Arabic? One solution can be found in a process called the Semantic Reverse Query Expander (SRQE). Based on components such as language ontologies, the SRQE process utilizes Java 2 Platform, Enterprise Edition (J2EE) (J2EE)[1] services that can take a term in one language (source language), expand the term conceptually, translate the expanded terms (into a target language), and perform a query on a targeted foreign language document set. Returns are translated into the language of the requestor. This Technical Report identifies an existing process and architecture used to query foreign language text files.

Technologies and ontologies (i.e. thesauri) for undertaking this kind of matching and expansion operation have been available for some time (e.g. the work of CYC Corp, Global WordNet, Global WordGrid). Valuable lessons have been learned about what such technologies can and cannot accomplish. This Technical Report does not discuss these pre-existing technologies, or describe the improvement or change that the proposed process presented might represent. A particular approach (theory and practice) with respect to the context of difficulties experienced in regard to multilingual equivalencies and translation are presented in Annex C of this Technical Report.

In Clause 3 of this Technical Report, an implementation of the SRQE process is described in a web environment to help clarify the architecture described in Clause 4. Annex A contains possible linkages to ISO/IEC JTC 1, SC 36 projects and future areas of study.

The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that the process described in this Technical Report may involve the use of patents. ISO and IEC take no position concerning the evidence, validity and scope of these patent rights.

The holders of these patent rights have assured ISO and the IEC that they are willing to grant a free of charge license to an unrestricted number of applicants on a worldwide, non-discriminatory basis and under other reasonable terms and conditions to make, use, and sell implementations of the process contained in this Technical Report. In this respect, the statements of the holders of these patent rights are registered with ISO and IEC. Information may be obtained from the companies listed below.

Raytheon Company
Phillip Berestecki
Intellectual Property and Licensing
870 Winter Street
Waltham, Massachusetts 02451-1449
USA

NOTE 1   This Technical Report refers to one particular process or approach for performing reverse semantic queries; there are other approaches and processes that could be developed for these same purposes.

NOTE 2   The process is not dependent on particular database software, protocols, or data sets. Specific components used in the process are an implementation decision.

---

1)   A widely used platform for server programming in the Java programming language.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

Information technology — System Process and Architecture for Multilingual Semantic Reverse Query Expansion

1   Scope

This Technical Report identifies an example of a system-based process to index, query, translate, and manage components used in querying and translating documents in multiple foreign languages, enabling learners in learning, education, and training areas to effectively find and share documents on a global scale.

2   Terms and definitions

For the purposes of this document, the following terms and definitions apply.

NOTE   For this Technical Report, the following terms and definitions are not considered to be normative. They are informative, and apply only within the context of this Technical Report.

2.1
coordinate term
words that have the same hypernym

EXAMPLE   Boat, yacht, and shrimper, all have the same hypernym, ship.

NOTE   Adapted from ISO 1087-1:2000, definition 3.2.19.

2.2
entity extraction
process that seeks to locate, classify, and tag atomic elements in text into predefined categories

EXAMPLE   Names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

2.3
hypernym
superordinate concept
word that is more generic or broad than another given word

NOTE 1   Another term for a hypernym is a superordinate concept.

NOTE 2   Adapted from ISO 1087-1:2000, definition 3.2.13.

2.4
hyponym
subordinate concept
word that is more specific than another given term

NOTE 1   Another term for hyponym is a subordinate concept.

NOTE 2   Adapted from ISO 1087-1:2000, definition 3.2.14.

2.5
Java servlet
Java programming language objects that dynamically process requests and construct responses

2.6
meronym
constituent part of, or a member of, something

EXAMPLE   "Winchester Cathedral" is a meronym of "Church of England".

2.7
nominalization
use of a verb or an adjective as a noun with or without morphological transformation, so that the word can now act as the head of a noun phrase

2.8
word sense
⟨linguistics⟩ one of the meanings of a word

NOTE   A dictionary may have over 50 different meanings of the word "play", with each of these having a different meaning based on the context of the word usage in a sentence.

EXAMPLE   We went to see the **play** *Romeo and Juliet* at the theater.

                The children went out to **play** in the park.

3   Example SRQE Implementation

This clause provides an implemented example of the SRQE process. The simplistic example of the SRQE implementation illustrates a sequence of actions between a learner and the system. The example provided is based on a possible learning assignment made to a learner in producing a report on trucks using foreign language resources. A learner wanting international information in producing a report on trucks could use the system to gather international information related to trucks for possible inclusion in the report. The system can perform a cross lingual query on documents in a number of languages, and translate the documents into the learner's native language. The learner can optionally access a map of the locations listed in the text files returned by the query for improved comprehension of where the article originates from, or the location of where the subject of the article can be found.

In this example, the SRQE process utilizes a user interface, interacting with Java servlets to provide a web based User Interface (UI). The SRQE process flow is illustrated below in Figure 1. The SRQE process is a human-machine interactive process to perform cross-lingual queries. Human inputs are shown in the white box above the green arrow. Java servlet functions are shown below the green arrow. A UML diagram showing the Java servlets in a green box are linked to functions in the white box.
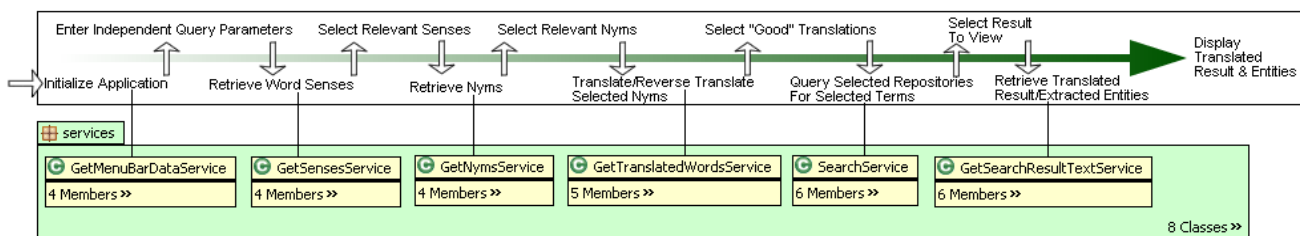

Figure 1 — SRQE Process Flow

3.1 Initialization of the User Interface

The user interface application is accessed by URL. The GetMenuBarDataService initializes the user interface application providing the repositories available and in what languages the repositories are in. The user interface application is shown in Figure 2. The GetMenuBarDataService builds the menus and menu choices in the user interface. This includes 1. Enter word to be translated, 2. Select Class, 3. Select Language(s), and 4. Select Sources, 5. Execute Query, the results section (blank), and the Map.
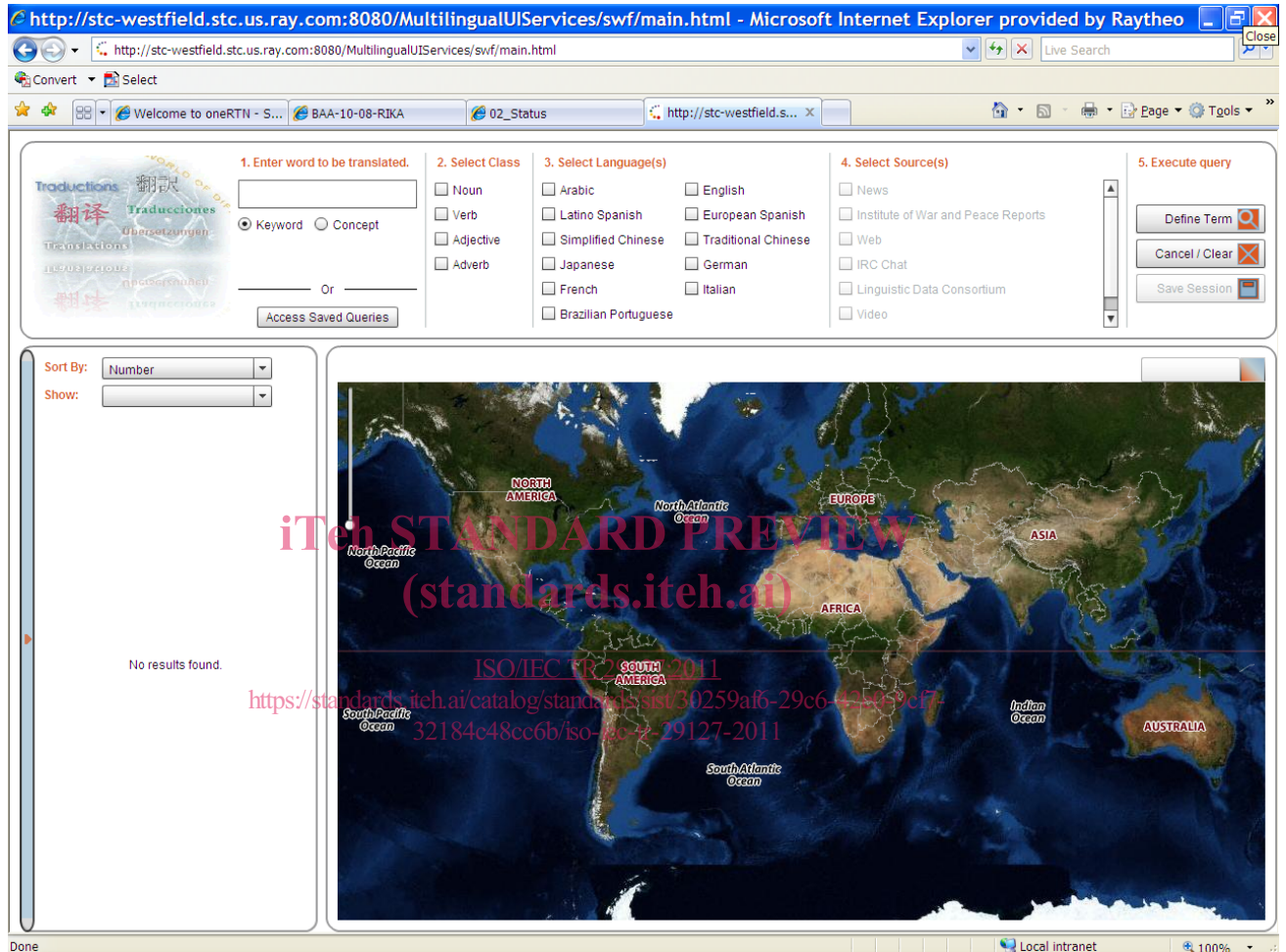
Figure 2 — Initialized User Interface

NOTE 1   In this example, an Adobe Flash plug-in is required for the use of the SRQE in a browser. The user interface described in the example is based on Adobe Flex.

NOTE 2   The original example uses CaMel CaSe format in this instance and the instances that follow. This formatting is retained for this reason.

3.2   Select Query Parameters

The learner enters a word in "1. Enter word to be translated". The learner selects a class of the word in "2. Select Class". The learner selects a language in "3. Select Language(s)". The learner selects a source in "4. Select Source(s)". In this example, the learner enters Truck in "1. Enter word to be translated". The learner clicks on Noun in "2. Select Class". The learner selects Arabic in "3. Select Language(s)". The learner selects Linguistic Data Consortium as the source in "4. Select Sources". Figure 3 shows menu items 1 through 4 filled in by the learner.
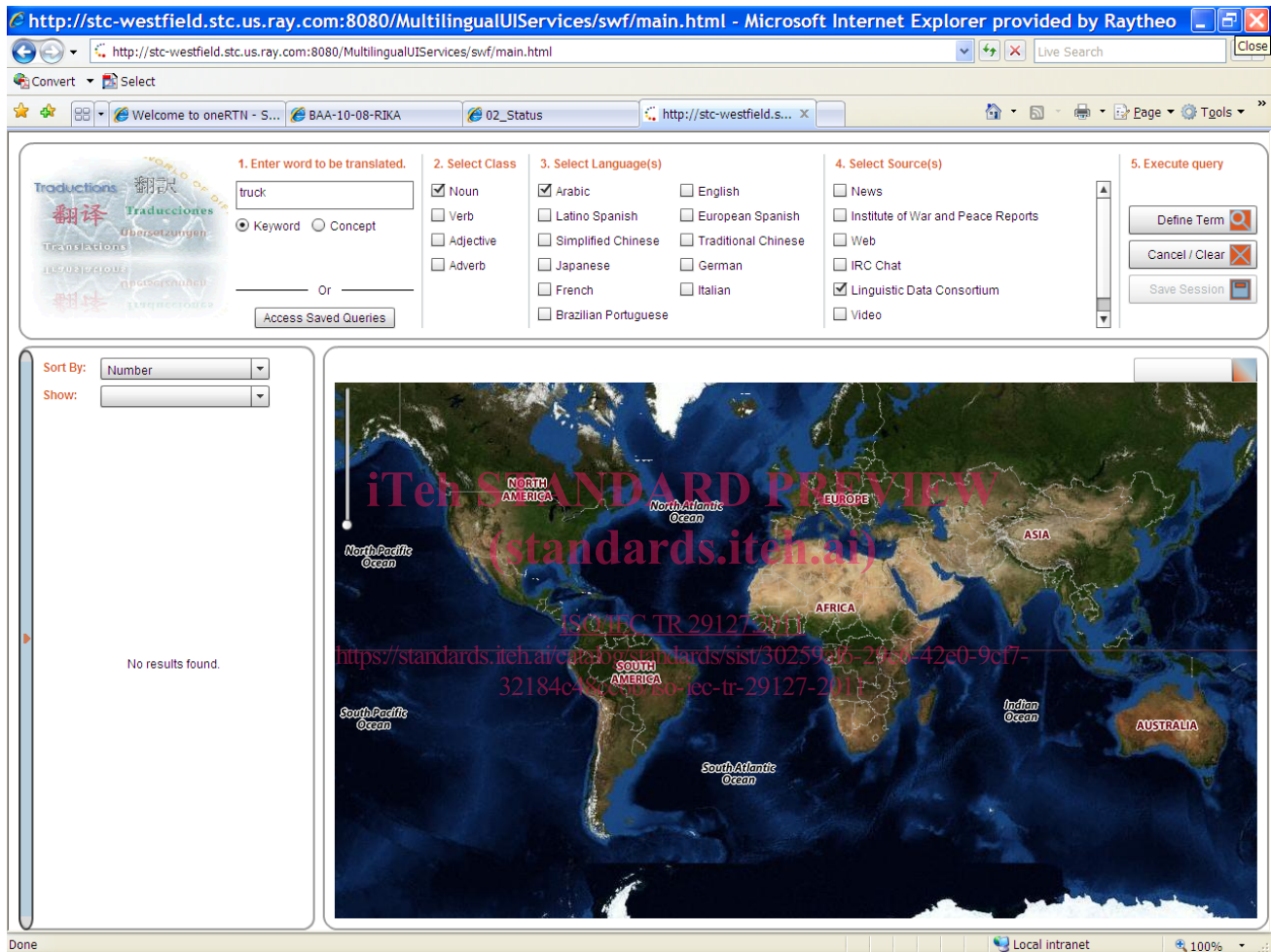
Figure 3 — Learner Inputs Menu Items 1-4

The learner selects Define Term in Item 5 "Execute query" menu. The GetSensesService retrieves the noun word senses for Truck, in this example, word senses are retrieved by GetSensesService from WordNet utilizing the Java WordNet Library (JWNL). The GetSensesService returns noun word senses for truck. The word senses for Truck is displayed in the return section of the interface. Figure 4 shows the noun word senses for Truck returned by the GetSensesService.

Figure 4 — Word Senses for Truck

3.3   Select Word Senses

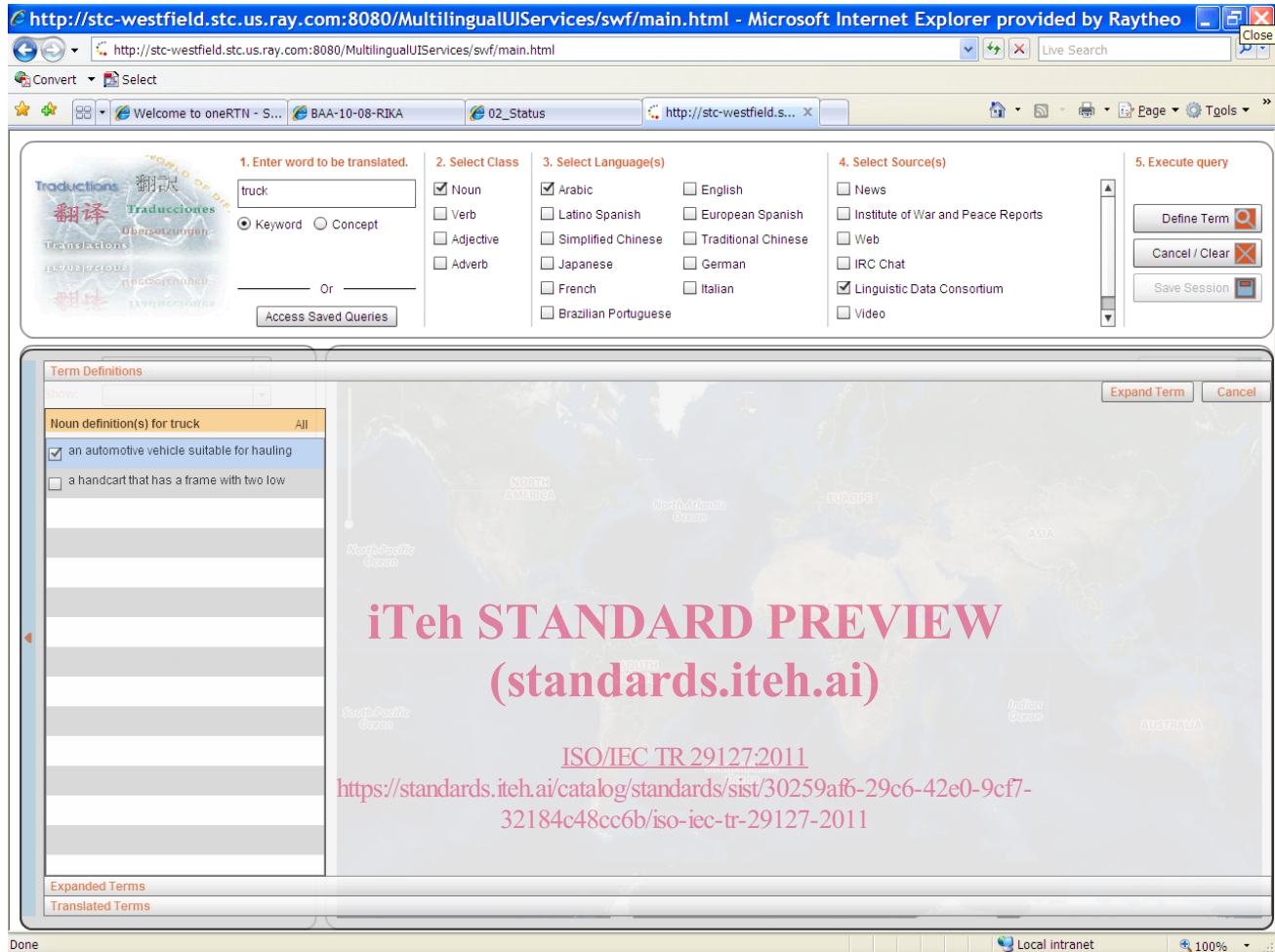The learner selects the relevant word sense for Truck. Figure 5 shows the word sense selected by the learner "an automotive vehicle suitable for hauling".

Figure 5 — Word Sense Selected