

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*

Reference number
ISO/IEC TR 29127:2011(E)





COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2011

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Terms and definitions	1
3 Example SRQE Implementation	2
3.1 Initialization of the User Interface	3
3.2 Select Query Parameters	4
3.3 Select Word Senses	6
3.4 Selecting and Translating Appropriate Terms	8
3.5 Selecting Appropriate Translations and Executing the Query	10
3.6 Query Returns	11
4 Components and Architecture of the SRQE Process	14
4.1 SRQE Process Flow	15
4.2 Repositories	15
4.3 Terms and Results	16
4.4 Translators	16
4.5 Entity Extraction	17
4.6 Terminology for Query Searches	17
Annex A (informative) Potential Linkage to Current and Future ISO/IEC JTC1 SC 36 Technology Areas	18
Annex B (informative) Patent Declaration Form for SRQE Process	20
Annex C (informative) Summary on the Issue of Language Equivalencies	22
Bibliography	30

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.

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.