InterSystems IRIS Basics: SQL Search

Version 2023.1
2023-05-26

InterSystems Corporation 1 Memorial Drive Cambridge MA 02142 www.intersystems.com

InterSystems IRIS Basics: SQL Search
InterSystems IRIS Data Platform Version 2023.1 2023-05-26
Copyright © 2023 InterSystems Corporation
All rights reserved.

InterSystems®, InterSystems IRIS®, IntegratedML®, InterSystems HealthShare®, InterSystems HealthShare Care Community®, HealthShare Unified Care Record®, InterSystems Caché®, and InterSystems Ensemble® are registered trademarks of InterSystems Corporation.InterSystems IRIS for Health™ and HealthShare® CMS Solution Pack™ are trademarks of InterSystems Corporation.

All other brand or product names used herein are trademarks or registered trademarks of their respective companies or organizations.

This document contains trade secret and confidential information which is the property of InterSystems Corporation, One Memorial Drive, Cambridge, MA 02142, or its affiliates, and is furnished for the sole purpose of the operation and maintenance of the products of InterSystems Corporation. No part of this publication is to be used for any other purpose, and this publication is not to be reproduced, copied, disclosed, transmitted, stored in a retrieval system or translated into any human or computer language, in any form, by any means, in whole or in part, without the express prior written consent of InterSystems Corporation.

The copying, use and disposition of this document and the software programs described herein is prohibited except to the limited extent set forth in the standard software license agreement(s) of InterSystems Corporation covering such programs and related documentation. InterSystems Corporation makes no representations and warranties concerning such software programs other than those set forth in such standard software license agreement(s). In addition, the liability of InterSystems Corporation for any losses or damages relating to or arising out of the use of such software programs is limited in the manner set forth in such standard software license agreement(s).

THE FOREGOING IS A GENERAL SUMMARY OF THE RESTRICTIONS AND LIMITATIONS IMPOSED BY INTERSYSTEMS CORPORATION ON THE USE OF, AND LIABILITY ARISING FROM, ITS COMPUTER SOFTWARE. FOR COMPLETE INFORMATION REFERENCE SHOULD BE MADE TO THE STANDARD SOFTWARE LICENSE AGREEMENT(S) OF INTERSYSTEMS CORPORATION, COPIES OF WHICH WILL BE MADE AVAILABLE UPON REQUEST.

InterSystems Corporation disclaims responsibility for errors which may appear in this document, and it reserves the right, in its sole discretion and without notice, to make substitutions and modifications in the products and practices described in this document.

For Support questions about any InterSystems products, contact:

InterSystems Worldwide Response Center (WRC)
Tel: +1-617-621-0700
Tel: +44 (0) 844 854 2917
Email: support@InterSystems.com

Table of Contents

InterSystems IRIS Basics: SQL Search 1
1 Why SQL Search is Important 1
2 How InterSystems IRIS Implements SQL Search 2
3 Learn More About SQL Search 2

InterSystems IRIS Basics: SQL Search

This article introduces you to InterSystems IRIS® data platform support for SQL text search, which provides semantic context searching of unstructured text data in a variety of languages. It covers the following topics:

- Why SQL Search is Important
- How InterSystems IRIS Implements SQL Search
- Learn More About SQL Search

To use SQL Search in a hands-on online exercise, see Searching Unstructured Text Data. For the full documentation on SQL Search, see the Using SQL Search.

A related, but separate, tool for handling unstructured texts is Natural Language Processing (NLP). SQL Search presupposes that you know what you are looking for. NLP text analysis allows you to analyze the contents of texts with no prior knowledge of the text contents.

1 Why SQL Search is Important

The ability to rapidly search unstructured text data is fundamental to accessing the content of the huge volume of text commonly stored by many companies and institutions. Any search facility for such data must have the following functionality:

- Fast search: InterSystems IRIS SQL Search can rapidly search large quantities of data because it is searching a generated optimized index to the data, rather than sequentially searching the data itself.
- Word-aware search: SQL Search is not a string search, it is a search based on semantic structures in the text. The most basic semantic structure for SQL Search is the word. This reduces the number of false positives that result when a string search finds a string embedded in another word, or when a string bridges two words.
- Entity-aware search: SQL Search takes into account multiple words that are grouped by semantic relationship to form entities. It can thus search for multiple words in a specified order (a positional phrase), words appearing within a specific proximity to each other (regardless of sequence), and words found at the beginning or at the end of an entity. This enables you to narrow a search to a word (or phrase) found in a specified context of other words.
- Language-aware search: identifying semantic relationships between words is language-specific. SQL Search contains semantic rules (language models) for ten natural languages. It also provides support for other languages. It does not require the creation or association of dictionaries or ontologies.
- Pattern matching: SQL Search provides both wildcard matching and regular expression (RegEx) matching to match character patterns.
- Fuzzy matching: SQL Search provides fuzzy search for near-matches that take into account a calculated degree of variation from the search string. This enables matching of spelling errors, among other things.
- Derived matching: SQL Search can use decompounding to match root words and component words. SQL Search can use synonym tables to match synonym words and phrases.

2 How InterSystems IRIS Implements SQL Search

SQL Search can search text data found in a column in an SQL table. In order to do this, you must create an SQL Search index for the column containing the text data. InterSystems implements a table column as a property in a persistent class.

There are three levels of index available, each supporting additional features as well as all of the features of the lower levels: Basic, Semantic, and Analytic:

- *Basic* supports word search and positional phrase search, including the use of wildcards, ranges between words in a phrase, regular expression (RegEx) matching, and co-occurrence search.
- *Semantic* supports all of the Basic functionality, and also supports InterSystems IRIS Natural Language Processing (NLP) entities. It can search for entities, and words or phrases that begin an entity or end an entity. It recognizes NLP attributes, such as negation.
- *Analytic* supports all of the Semantic functionality, and also supports NLP paths. It can also search based on NLP dominance and proximity scores.

Populating the index. Like all SQL indices, you can either build the index directly after the table has been populated with data, or have SQL automatically build the index entries as you insert records into an empty table. In either case, SQL automatically updates this index as part of subsequent insert, update, or delete operations.

You perform an SQL Search you write a SELECT query in which the WITH clause contains `%ID %FIND search_index()` syntax. The **search_index()** function parameters include the name of the SQL Search index and a search string. This search string can include wildcard, positional phrase, and entity syntax characters. The search string can also include AND, OR, and NOT logical operators.

3 Learn More About SQL Search

InterSystems has other resources to help you learn more about SQL Search, including:

- Searching Unstructured Text Data
- Using InterSystems SQL Search