Capgemini

SAP HANA

Lesson Name: FUZZY Search

OurUniversity
EFMD CLIP ACCREDITED

Contents

- Introduction to Fuzzy Search
- Examples of Fuzzy Search
- Configuration in CDS View

Lesson Objectives

After completing this lesson, participants will be able to -

- Know the meaning of Fuzzy Search and How to Implement In HANA Data base.
- Configure Fuzzy search logic.

Introduction to Fuzzy Search

What is Fuzzy search

- Fuzzy Search is a fast and fault-tolerant search feature for SAP HANA.

Definition:

- The term ”fault-tolerant search” means that a database query returns records even if the search term (the user input) contains additional or missing characters or other types of spelling errorenterprise.
- Fuzzy search can be used in various applications, for example:
- Search for documents on ‘Driethanolamyn’ and find all documents that contain the term ‘Triethanolamine’.
- Fault-tolerant search in structured database content: Search for a product called ‘coffe krisp biscuit’ and find ‘Toffee Crisp Biscuits’
- Fault-tolerant check for duplicate records: Before creating a new customer record in a CRM system, search for similar customer records and verify that there are no duplicates already stored in the system.

Introduction to Fuzzy Search

- So in short it can find values in big data sets without actually hitting the search term exactly right.
- The fuzzy search algorithm calculates a fuzzy score for each string comparison. The higher the score, the more similar the strings are. A score of 1.0 means the strings are identical. A score of 0.0 means the strings have nothing in common.
- You can request the score in the SELECT statement by using the SCORE() function.
- You can sort the results of a query by score in descending order to get the best records first the best record is the record that is most similar to the user input). When a fuzzy search of multiple columns is used in a SELECT statement, the score is returned as an average of the scores of all columns used. So not only does it find a "fault tolerant" match, it also puts a score behind it:
- For example, when searching with 'SAP', a record like 'SAP Deutschland AG & Co. KG' gets a high score, because the term 'SAP' exists in both texts. A record like 'SAPPHIRE NOWOrlando' gets a lower score, because 'SAP' is only a part of the longer term 'SAPPHIRE' (3 of8 characters match)

Examples of Fuzzy Search

Example: 1 (Twitter data)

- Twitter data from millions of tweets
- This is a download of Twitter data from March 2006 to November 2009
- The data set consists of “tokens,” which are hashtags (#data), URLs, or emoticons (Twitter smileys or other “faces” created using keyboard characters)
- The data comes from analysis on the full set of tweets during that time period, which is 35 million users, over 500 million tweets, and more than 1 billion relationships between users.
- When doing a fuzzy search, The table needs to be structured in a certain way. This can easily be done by using SQL to create your table. This can be done by the following command:

Examples of Fuzzy Search

CREATE COLUMN TABLE "S0007457730"."TWITTERINFOCHIMPS3"
("TOKEN_TYPE" **VARCHAR**(10)**NOT NULL** ,
"YEAR_MONTH" **VARCHAR**(6),
"COUNT" **INTEGER** CS_INT,
"TOKEN_TEXT" SHORTTEXT(200) **LANGUAGE** DETECTION ('EN')
PHRASE **INDEX** RATIO 0.200000 FUZZY
SEARCH **INDEX ON** SEARCH **ONLY ON** FAST PREPROCESS **ON**)

In general fuzzy searches can be performed on:

- TEXT
- SHORTTEXT
- VARCHAR, NVARCHAR
- DATE
- All data types with a full-text index

Examples of Fuzzy Search

If we choose SHORTTEXT as Text type fuzzy searches can be the most sophisticated. Note that the 'FUZZY SEARCH INDEX' structure will improve performance as it increases the memory of the loaded table. Set it in advance as you cannot modify it later.

DCD (S0007457730) hanasvr-04.sapdevcenter.com 00

Table Name:

TWITTERINFOCHIMPS3

Columns | Indexes | Further Properties | Runtime Information

	Name	SQL Data Type	Dim	Column Store Data Type	Key	Not Null	Default	Comment
1	TOKEN_TYPE	VARCHAR	10	STRING		X		
2	YEAR_MONTH	VARCHAR	6	STRING				
3	COUNT	INTEGER		INT				
4	TOKEN_TEXT	SHORTTEXT	200	STRING				

Examples of Fuzzy Search

Example 2: (Give all records containing ‘SAP’ as part of the hash tag)

```
SELECT SCORE() AS score, * FROM TWITTERINFOCHIMPS3
WHERE CONTAINS(TOKEN_TEXT,’SAP’,
FUZZY(0.7,’textSearch=compare,bestMatchingTokenWeight=0.7’))
 AND TOKEN_TYPE = ‘hashtag’
 ORDER BY score DESC;
```

The dataset comes back with over 800 records and my highest ranked scored values are:

Continued in Next Page.

Examples of Fuzzy Search

The Results continued from the last page:

SCORE	TOKEN_TYPE	TOKEN_TEXT
1,00	hashtag	sap
0,94	hashtag	sap_wt
0,94	hashtag	sap_inovation
0,94	hashtag	sap_innovation
0,94	hashtag	sap_bw

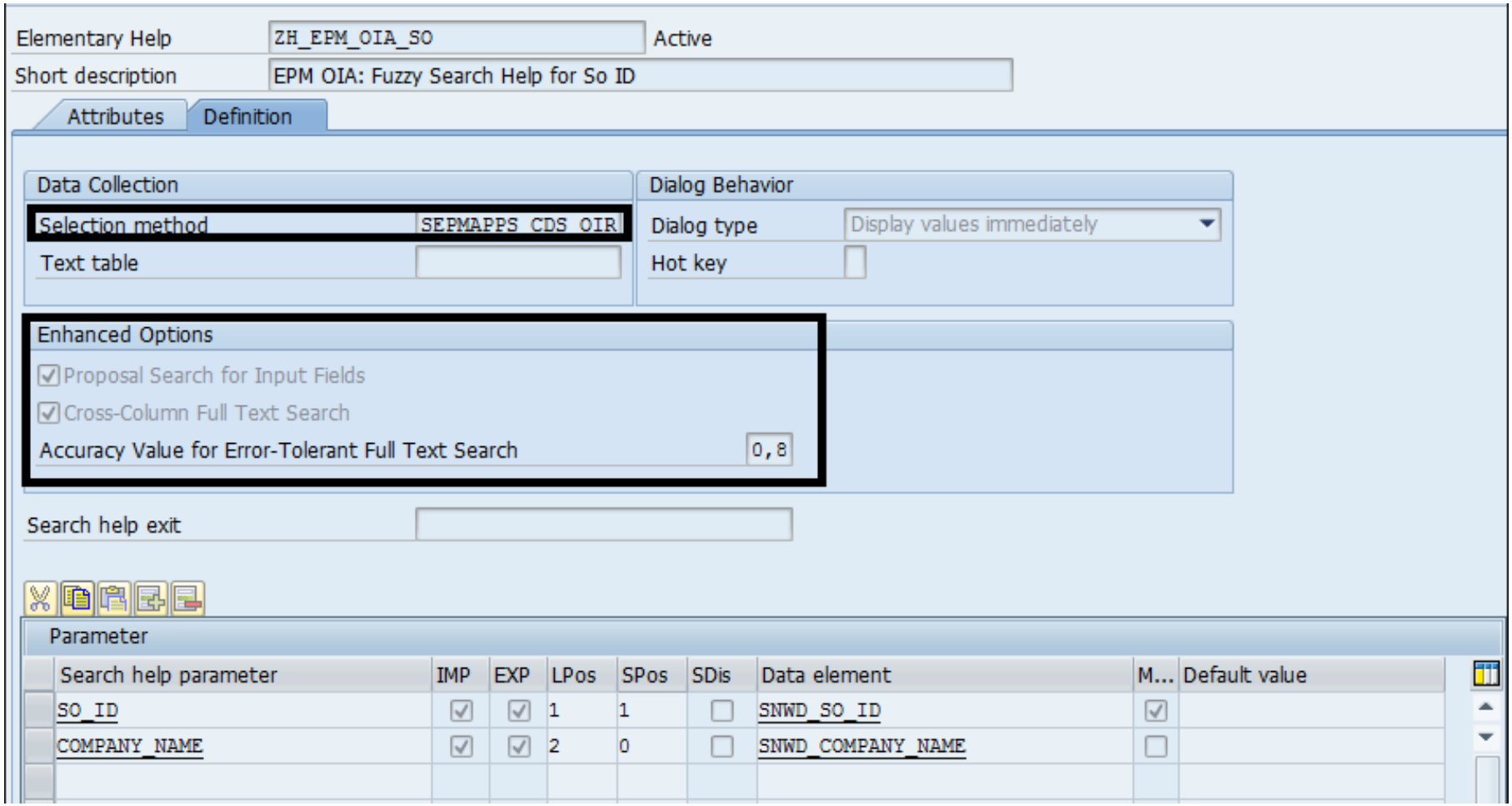
Configuration in CDS View

Fuzzy search logic is implemented in CDS Views.

The steps to configure Fuzzy search logic is as below.

- 1) Create a Search Help [***ZH_EPM_OIA_SO***] using the ***SE11*** transaction. Make sure it is an Elementary Search Help.
- 2) Enter the name of the CDS view name [***SEPMAPPS_CDS_OIR***] in "Selection method" input field.
- 3) Use the "***Enhanced Options***" section to configure the fuzzy search options [Refer screenshot given in Next step 4].
- 4) Create Search help parameters such as ***SO_ID*** and ***COMPANY_NAME*** and mark them as importing parameter, if you want to search for both *company_name* and *so_id*. You can also mark both the parameters as exporting parameters, if you want to display both, *company_name* and *so_id*, as output of the fuzzy search as shown in the snapshot below.

Configuration in CDS View



Configuration in CDS View

5) Currently, the ***SE11*** transaction does not support enhanced options in the test mode. You need to test them in a normal screen or by using a report with the type-ahead API.

6) **Fuzzy Search** is a fast and fault tolerant search and finds strings that match a pattern approximately(rather than exactly) means it returns records even if the search term contains additional or missing characters or other types of spelling error.

7) Fuzzy search algorithm calculates a fuzzy score for each string comparison.

8) We can call Fuzzy Search by using CONTAINS() function with FUZZY() option in WHERE clause of a SELECT Statement such as:

Configuration in CDS View

Syntax for Fuzzy search as below:

```
SELECT * FROM DOCUMENTS
WHERE
CONTAINS  (doc_content, 'ox', FUZZY(0.7))
```