RINGGOLD

Identify Database REST API Service

Search Tools 1.3 Operation and Usage

Ver 2019-03-04
Status: Released

Modified by: bill.haase@ringgold.com

Search Tools Library 1.3 (aka STL1.3)

Summary: STL1.3 is a php+mysql library written to search the Identify database utilizing both intuitive simple search line and more advanced search string methods.

Note: In this document, token and keyword are interchangeable.

Vocabulary

1) Full text (aka FT) an index of words/tokens found in the indexed element/field.

STL1.3

1) is used Ringgold's REST API (API) ver 2.6+.
    a) REST API <= 2.5.x call STL1.2 to search
2) Utilizes MySQL full text indexes of our Identify database.
    a) Full text indexing creates an index of keywords/tokens.
    b) All non-alphanumeric characters except underscore and apostrophe use to define breaks between tokens.
    c) We now have two indexes for names and urls to improve relevance scoring between names usually stored in two different name tables.
    d) Several Organization elements have full text (FT) indexes:
        i) Org.names (org.comp_name and alt.names)
        ii) Org.urls (org.url + alt.urls)
    e) Our full text indexes have limitation on what words are indexed. Short, long and very common words (called Stopwords) are not indexed and can't be searched in the FT indexes. They can be used in other searches explained below.
        i) Short words are 1-2 characters (less than 3).
        ii) Long words are greater than 84 characters long.
        iii) There are 36 pre-defined Stop Words:

(1) "a", "about", "an", "are", "as", "at", "be", "by", "com", "de", "en", "for", "from", "how", "i", "in", "is", "it", "la", "of", "on", "or", "that", "the", "this", "to", "was", "what", "when", "where", "who", "will", "with", "und", "the", "www"

3) Search Process
   a) Calls to the API require a mode parameter to control the indexes searched;
      i) 'name' performs a name search
      ii) 'url' performs a url search
      iii) 'name_url' searches both names and urls
      iv) 'isni'= performs an isni search ** NEW
   b) All searches are case-insensitive
   c) Search string is processed to determine FT searchable tokens.
      i) All unique tokens in the search string separated by one of these characters: space, comma, semi-colon, colon, forward slash, are compared to the stopword lengths and stopword list and added to the searchable tokens if not a stopword.
   d) FT searching allows modifier characters immediately in front of each token:
      plus, minus, tilde, greater than sign, less than sign.
      These tell the FT search to require or exclude some of the tokens.
      i) By default, the token list is a 'OR' list, the occurrence of any of the words will report a search hit.
      ii) Adding modifier plus '+' in front of tokens
         (1) Changes the FT search to an 'AND' search
         (2) Requiring all plused tokens to appear in candidate field.
         (3) E.g. search=+universistat +koln, will only yield orgs with both tokens.
      iii) Adding modifier minus/hyphen '-' in front of tokens
         (1) Changes the search to exclude records with that token.
         (2) E.g. search=+portland -university, will only include orgs with portland but not if they also contain university.
      iv) Adding modifier tilde '~'
         (1) Inverts the relevance score for that token.
         (2) Subtracts that token value from the total score for org.
      v) Add modifier greater than sign '>'
         (1) Increases the token's value but doesn't require it.
      vi) Add modifier less than sign '<'
         (1) Decreases the token's value but doesn't require it
      vii) Wildcard '*'
         (1) Can only be added to end of a token
         (2) E.g. +univ* will require at least one token that begins with univ
         (3) E.g. -univ* will exclude any orgs where name includes token starting with univ
   e) Search steps
      i) Exact match search, where value equals "search string"

ii) Begins with search, where value begins with “search string”
iii) Phrase search, where value contains “search string”
iv) keyword search, where value contains search tokens

4) Search API Params (passed as lowercase)
   a) ‘q’ - (required)
      i) phrase or tokens with optional FT modifiers
      ii) isni
         (1) (opt. w/spaces or w/hyphens).
            (a) eg 0000 0000 0000 0000
            (b) eg 0000-0000-0000-0000
            (c) eg 0000000000000000
         (2) (opt wo/leading zeros)
            (a) 2345678
            (b) 234-5678
            (c) 234 5678
            (d) (all treated as 0000000002345678)
   b) ‘mode’ - (required)
      i) name
      ii) url
      iii) name_url
      iv) isni
   c) ‘city’ (aka place.name)
   d) ‘state’ (aka place.aal1)
   e) ‘country’ (aka place.country)
   f) ‘postcode’
   g) ‘limit’
   h) ‘offset’
   i) ‘ofr’
      i) default (when not specified) include any org
      ii) 0 = only include orgs with NO OFR
      iii) 1 = only include orgs with an OFR
   j) ‘out’’
      i) 0 = (default when not supplied) = short results / only ringgold_id
      ii) 1 = brief results, ringgold_id, max_score, org_hits, hit_values
      iii) 2 or true = complete result / all org elements
   k) ‘pretty’ - json pretty print
      i) default = off
      ii) 1 or true = pretty print the json response returned to calling client