Parameter descriptions for Distingo Context

Summary: This document describes the parameters of Distingo Context in XML mode and in programming mode. The first term is the name of the parameter in XML, followed by the accessors for writing in programming mode.

MODE (setModeDensity – if the parameter is set to 1, the mode is DENSITY): if the mode is DENSITY, Distingo Context does not deduce additional semantics from existing phrases and it does not merge phrases having the same subject. The result of the matching between every phrase in the query text and every phrase in the target text is returned.

If the mode is PRESENCE, Distingo Context does its best to match the query in the document text by deducing additional semantics and merging phrases having the same subject. The result of the matching between every phrase in the query text and the best matched phrase in the target text is returned.

DISTADJNOUN (setDistAdjNoun) : the maximum distance in the ontology between two nouns or two adjectives, while still allowing Distingo to consider them similar. (The value must be between 1 and 6, and the default value is 3.)

DISTVERB (setDistVerb) : the maximum distance in the ontology between two verbs, while still allowing Distingo to consider them similar. (The value must be between 1 and 6, and the default value is 3.)

> **First Note**: the similarity (a value between 0 and 100) between two words is calculated in function of its **distance** (see DISTADJNOUN AND DISTVERB). If the distance is 0, the similarity is 100; if the distance is greater than the maximum distance allowed (in DISTVERB or DISTADJNOUN), the similarity is 0. In all other cases, the value will be between these extremes.

> **Second Note**: the greater the value of DISTADJNOUN or DISTVERB, the longer the calculation time – and thus the slower the generation of results.

WEIGHTSUBJECT (setWeightSubject) : in the calculation of the average similarity of the corresponding elements between two concepts (MATCHINGSIMILARITY), WEIGHTSUBJECT is the weight given to the average of the similarity between elements of the subject (nouns and adjectives identified as the subject) (always a value between 1 and 100, with a default value of: 40).

WEIGHTVERB (setWeightVerb) : in the calculation of the average similarity of the corresponding elements between two concepts (MATCHINGSIMILARITY), WEIGHTVERB is the weight given to the average of the similarity between elements of the verb (always a value between 1 and 100, with a default value of: 35).

WEIGHTPREDICATE (setWeightPredicate) : in the calculation of the average similarity of the corresponding elements between two concepts (MATCHINGSIMILARITY), WEIGHTPREDICATE is the weight given to the average of the similarity between elements of the predicate/object (nouns and adjectives identified as the subject) (always a value between 1 and 100, with a default value of: 25).

Third Note: the total of WEIGHTSUBJECT, WEIGHTVERB and WEIGHTPREDICATE must equal 100.

For two concepts to be identified as linked (PRESENCE mode) or for two phrases to be identified as linked (DENSITY mode), the following 3 threshold values much be reached:

SUBJECTTHRESHOLD (setSubjectThreshold) : minimum similarity threshold of the subject (noun) between a subject in the query and a subject in the target text; the similarity is calculated as explained in **First Note**, above (with a value between 1 and 100; default value of 20).

Fourth Note: note that a high threshold (more than 50) will require that elements be very close in the ontology.

MATCHINGTHRESHOLD (setMatchingThreshold) : minimum threshold of the percentage of elements (adjectives, verbs, nouns) that have a similar element in the target text. For a value of 60%, this means that 60% of the elements find a match in the target text (leaving a remainder of 40%). (The value must be between 1 and 100; the default value is 60).

MATCHINGSIMILIARITYTHRESHOLD (setMatchingSimiliarityThreshold) : minimum threshold for the average similarity of elements for which there exists a similar element with the target text. This similarity is calculated by using WEIGHTSUBJECT, WEIGHTVERB and WEIGHTPREDICATE. (The value must be between 1 and 100; the default value is 20).

Fifth Note: note that, as for SUBJECTTHRESHOLD, a high threshold (more than 50) will require that elements be very close in the ontology.

USEADJ (setUseAdjective – if the parameter is set to 1, use adjectives) : if the value is TRUE, Distinguo Context takes into account the adjectives present in both query and target text otherwise it does not.

PREDICATEMATCHING (setPredMatchingTight – if the parameter is set to 1, the matching of predicate is tight) : if the value is TIGHT, predicates in the target text must match tightly the query text otherwise they don't have to.

For example, let's suppose the query text is: "The man eats a big sandwich" and the target text is "The man eats a sandwich".

If the value is TIGHT, there is not matching since the query contains the fact the man doesn't any sandwich but a big sandwich.

If the value is not TIGHT, the matching is done.

Please note that if the target text is "The man eats a big sandwich" (or "The man eats a large sandwich"), the matching is done whatever is the value of PREDICATEMATCHING.

SUGGESTDEEPMODE (setSuggestDeepMode – if the parameter is set to 1, suggestion mode is deep): if the value of SUGGESTDEEPMODE is TRUE, Distinguo uses a deeper calculation of suggestions in order to correct words that are misspelled in the text. This mode is slower, but more effective.

DISTWORD (setDistWord) : the maximum allowed distance in the ontology between two words in linguistic mode; beyond this distance, Distinguo will not consider words related. (The value must be between 1 and 6; the default value is 3).