Λημματοποιητής

Παρ, 16/05/2008 - 16:25 — webmaster

Περιγραφή

Λημματοποιητής (Lemmatizer) είναι το γλωσσικό εργαλείο που δέχεται ως είσοδο έναν οποιοδήποτε λεκτικό τύπο και επιστρέφει το λημματικό τύπο στον οποίο αντιστοιχεί, π.χ. για τον τύπο *κατέστη* επιστρέφει το λημματικό τύπο *καθιστώ*. Στην περίπτωση που ο δοθείς τύπος αντιστοιχεί σε περισσότερους του ενός λημματικούς τύπους, ο Λημματοποιητής τους επιστρέφει όλους, π.χ. για τον τύπο *απαντήσεις* επιστρέφει *απαντώ* και *απάντηση*. Η λειτουργία του Λημματοποιητή της Neurolingo βασίζεται σε ευρετήριο που περιέχει όλους τους λεκτικούς τύπους όλων των λημμάτων του Μορφολογικού Λεξικού [1].

Για μια γλώσσα με έντονη μορφολογία όπως η Νέα Ελληνική, ο Λημματοποιητής αποτελεί αναπόσπαστη λειτουργική συνιστώσα σε εφαρμογές ευρετηριασμού και αναζήτησης κειμενικής πληροφορίας. Για παράδειγμα, όταν ο χρήστης αναζητά κείμενα που περιέχουν τον όρο *υπολογιστές* , προφανώς θα ήθελε μέσα στα αποτελέσματα αναζήτησης να περιλαμβάνονται και κείμενα που περιέχουν τους τύπους *υπολογιστής*, *υπολογιστή* και *υπολογιστών*. Αυτό πρακτικά επιτυγχάνεται μέσω λημματοποίησης τόσο των όρων ευρετηριασμού των κειμένων όσο και των όρων που απαρτίζουν τα ερωτήματα αναζήτησης: κείμενα που περιέχουν τους τύπους *υπολογιστή*, *υπολογιστές* και *υπολογιστών* θα ευρετηριαστούν και βάσει του λημματικού τύπου *υπολογιστής*· ένα ερώτημα που περιέχει τον τύπο *υπολογιστές* θα διευρυνθεί ώστε να περιέχει και το λημματικό τύπο *υπολογιστής*. Έτσι, ένα κείμενο που περιέχει τον τύπο *υπολογιστών* μπορεί πλέον να συσχετιστεί με ένα ερώτημα που περιέχει τον τύπο *υπολογιστές* μέσω του κοινού λημματικού τύπου *υπολογιστής*.

Δοκιμάστε online το Λημματοποιητή. [2]

Εφαρμογές

Ο Λημματοποιητής της Neurolingo έχει ενσωματωθεί στα παρακάτω συστήματα ευρετηριασμού και αναζήτησης κειμένων:

- Apache Lucene [3]. Η λειτουργικότητα του Λημματοποιητή παρέχεται μέσω απογόνου της Java κλάσης org.apache.lucene.analysis.Analyzer.
- Oracle Text [4]. Η λειτουργικότητα του Λημματοποιητή παρέχεται μέσω stored procedures (USER_LEXER preference).
- Microsoft Indexing Service [5] / SQL Server Full-Text Search [6]. Η λειτουργικότητα του Λημματοποιητή παρέχεται μέσω υλοποίησης του IStemmer COM interface.

Source URL: http://www.neurolingo.gr/el/technology/application_tools/lemmatizer.jsp

Links:
[1] http://www.neurolingo.gr/el/technology/lexica/morpholexicon.jsp
[2] http://www.neurolingo.gr/el/online_tools/lexiscope.htm
[3] http://lucene.apache.org
[4] http://www.oracle.com/technology/products/text/index.html
[5] http://msdn2.microsoft.com/en-us/library/bb187804.aspx
[6] http://msdn2.microsoft.com/en-us/library/ms345119.aspx

©2005-2010 Neurolingo E.E.

©2005-2010 Neurolingo E.E.