



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Εργαστήριο Τεχνολογίας Λογισμικού, Ιστού και
Ευφυών Συστημάτων

ΜΑΘΗΜΑ:	Υπολογιστική Λογική και Λογικός Προγραμματισμός
ΕΡΓΑΣΙΑ:	Υποχρεωτική
ΒΑΘΜΟΛΟΓΙΑ:	0 - 10
ΗΜΕΡΟΜΗΝΙΑ ΑΝΑΚΟΙΝΩΣΗΣ:	12/04/2016
ΗΜΕΡΟΜΗΝΙΑ ΠΑΡΑΔΟΣΗΣ:	03/06/2016
ΤΡΟΠΟΣ ΠΑΡΑΔΟΣΗΣ:	<u>Upload στο elearning.auth.gr</u>

Περιγραφή Άσκησης

Γενική περιγραφή προβλήματος: Έστω ένα πρόγραμμα – βοηθός ενός επιστήμονα Πληροφορικής ο οποίος επιθυμεί να καταθέσει μια εργασία σε ένα επιστημονικό συνέδριο. Το συνέδριο είναι χωρισμένο σε ειδικότερες «συνεδρίες» sessions. Κάθε session έχει ξεχωριστή θεματική, η οποία απεικονίζεται τόσο από τον γενικό τίτλο του session, όσο και από τα επιμέρους θεματικά αντικείμενα που περιγράφουν ειδικότερα επιστημονικά αντικείμενα τα οποία θα παρουσιαστούν στο session. Ένας επιστήμονας που θέλει να καταθέσει μία εργασία στο συνέδριο θα κληθεί να επιλέξει ένα από τα sessions με βάσει το αντικείμενο που πραγματεύεται η εργασία του. Το πρόγραμμα – βοηθός θα παίρνει ως είσοδο από τον χρήστη κάποιες λέξεις-κλειδιά (keywords) οι οποίες περιγράφουν το θέμα της εργασίας του και θα προτείνει σε φθίνουσα σειρά ταξινόμησης τα πιο σχετικά sessions μαζί με το αντίστοιχο σκορ σχετικότητας.

Ειδικότερη περιγραφή επίλυσης: Έστω ότι υπάρχει μια βάση δεδομένων (γεγονότα της Prolog) που αναπαριστά ένα σύνολο από sessions στα οποία θέλουμε να κάνουμε αναζήτηση σχετικότητας για την εργασία βάσει λέξεων-κλειδιών (keywords). Τα γεγονότα που αναπαριστούν τα sessions είναι της μορφής **session (Title, Topics)**, όπου **Title** είναι ο γενικός τίτλος του session και **Topics** είναι μία λίστα από «θέματα-αντικείμενα» τα οποία θα παρουσιαστούν στο session. Ακολουθεί ένα παράδειγμα ενός τέτοιου γεγονότος:

session ('Rules and Norms',

*['Methodologies for modeling regulations using both ontologies and rules',
'Defeasibility and norms - modeling rule exceptions and priority relations among rules',
'The relationship between rules and legal argumentation schemes',
'Rule language requirements for the isomorphic modeling of legislation',
'Rule based inference mechanism for legal reasoning',
'E-contracting and automated negotiations with rule-based declarative strategies']*) .

Στην διεύθυνση <https://elearning.auth.gr/mod/resource/view.php?id=281681> (στο elearning.auth.gr) υπάρχει το αρχείο με όλα τα γεγονότα που θα χρειαστείτε για να δοκιμάσετε τον κώδικά σας. Φυσικά αν θέλετε μπορείτε να το εμπλουτίσετε, χωρίς να είναι απαραίτητο.

Στην εργασία, θα υλοποιήσετε το κατηγορημα **query (ListOfKeywords)** το οποίο θα δέχεται μια λίστα από λέξεις κλειδιά **ListOfKeywords** και θα τυπώνει στην οθόνη μία λίστα από κείμενα

(τίτλους) σχετικά με τις λέξεις-κλειδιά του ερωτήματος (query). Η σχετικότητα ενός κειμένου καθορίζεται ως εξής:

- Για να είναι σχετικό ένα κείμενο, θα πρέπει οι λέξεις-κλειδιά να υπάρχουν μέσα σε κάποια από τις συμβολοσειρές των «θεμάτων» του κειμένου ή στο τίτλο του κειμένου.
- Κάθε κείμενο θα λαμβάνει ένα σκορ σχετικότητας ανάλογα με το πόσες λέξεις-κλειδιά του ερωτήματος περιέχει.
- Οι λέξεις-κλειδιά μπορούν να έχουν και βάρος, ενώ η έλλειψη βάρους αντιστοιχεί σε βάρος 1. Π.χ. [**exception-2,rule**] που αντιστοιχεί στην λίστα λέξεων-κλειδιών [**exception-2,rule-1**].
- Οι λέξεις-κλειδιά μπορούν να είναι και μικρές φράσεις π.χ. '**rule exception**'. Στην περίπτωση αυτή ως λέξεις-κλειδιά θεωρούνται και οι λέξεις από τις οποίες αποτελείται η φράση, με βάρος **W/N**, όπου **W** το βάρος της «φράσης» και **N** ο αριθμός των λέξεων της φράσης. Π.χ. [**'rule exception'**] που αντιστοιχεί στην λίστα λέξεων-κλειδιών [**'rule exception'-1,rule-0.5,exception-0.5**].
- Η συνάρτηση που υπολογίζει το σκορ σχετικότητας κάθε εγγράφου είναι η εξής:
 - ο Ψάχνω όλες τις λέξεις κλειδιά μέσα σε κάθε «θέμα» ενός session. Κάθε φορά που βρίσκω μια λέξη-κλειδί σε ένα «θέμα» το σκορ του «θέματος» αυξάνεται κατά **W**, όπου **W** το βάρος της λέξης-κλειδί. Π.χ. το «θέμα» *'Defeasibility and norms - modeling rule exceptions and priority relations among rules'* σε σχέση με τη λίστα λέξεων-κλειδιών [**'rule exception'-1,rule-0.5,exception-0.5**] λαμβάνει σκορ σχετικότητας 2.
 - ο Ο τίτλος του session βαθμολογείται με τον ίδιο ακριβώς τρόπο, αλλά η βαθμολογία του πολλαπλασιάζεται επί 2, γιατί θεωρείται πολύ πιο σημαντικό να βρίσκονται οι λέξεις-κλειδιά στον τίτλο του sessions.
 - ο Το σκορ σχετικότητας του session προκύπτει από τα σκορ σχετικότητας των «θεμάτων» και του τίτλου του sessions ως εξής: Επειδή θέλουμε να «επιδοτήσουμε» το «θέμα» (συμπεριλαμβανομένου και του τίτλου) με το μεγαλύτερο σκορ, δηλαδή με τις περισσότερες «ταιριασμένες» λέξεις-κλειδιά, πολλαπλασιάζουμε το σκορ του με 1000. Επειδή θέλουμε να μετράνε (δευτερευόντως) και τα σκορ των υπολοίπων «θεμάτων» προσθέτουμε στον παραπάνω αριθμό και το άθροισμα των σκορ όλων των «θεμάτων» (και του τίτλου). Π.χ. το sessions '*Rules and Norms*' παραπάνω, έχει τα παρακάτω σκορ για τον τίτλο και τα θέματά του:
 - '*Rules and Norms*': σκορ $0.5 \cdot 2 = 1$
 - '*Methodologies for modeling regulations using both ontologies and rules*': σκορ 0.5
 - '*Defeasibility and norms - modeling rule exceptions and priority relations among rules*': σκορ $0.5 + 0.5 + 1 = 2$
 - '*The relationship between rules and legal argumentation schemes*': σκορ 0.5
 - '*Rule language requirements for the isomorphic modeling of legislation*', σκορ 0.5
 - '*Rule based inference mechanism for legal reasoning*', σκορ 0.5
 - '*E-contracting and automated negotiations with rule-based declarative strategies*' σκορ 0.5

Συνεπώς το σκορ ολόκληρου του sessions θα είναι $1000 \cdot \max\{1, 0.5, 2, 0.5, 0.5, 0.5, 0.5\} + \text{sum}\{1, 0.5, 2, 0.5, 0.5, 0.5, 0.5\} = 2005.5$

Για κάθε ερώτημα που τίθεται από τον χρήστη θα βαθμολογούνται όλα τα sessions, θα ταξινομούνται με φθίνουσα σειρά σχετικότητας, και θα τυπώνονται οι τίτλοι τους, μαζί με το αντίστοιχο σκορ σχετικότητας.

Παραδείγματα χρήσης:

```
?- query(['rule exceptions']).
```

Session: Rules and Norms
Relevance = 2005.5
Session: RuleML-2010 Challenge
Relevance = 1005.5
Session: Rule-Based Distributed/Multi-Agent Systems
Relevance = 1005.5
Session: Rules and Inferencing
Relevance = 1005.0
Session: Rules and Uncertainty
Relevance = 1003.0
Session: Rule-based Event Processing and Reaction Rules
Relevance = 1003.0
Session: Rule Transformation and Extraction
Relevance = 1003.0
Session: General Introduction to Rules
Relevance = 1002.5
Session: Rules; Semantic Technology; and Cross-Industry Standards
Relevance = 1002.0
true.

?- query([exceptions-2,rule]).
Session: Rules and Norms
Relevance = 3010
Session: RuleML-2010 Challenge
Relevance = 2011
Session: Rule-Based Distributed/Multi-Agent Systems
Relevance = 2011
Session: Rules and Inferencing
Relevance = 2010
Session: Rules and Uncertainty
Relevance = 2006
Session: Rule-based Event Processing and Reaction Rules
Relevance = 2006
Session: Rule Transformation and Extraction
Relevance = 2006
Session: General Introduction to Rules
Relevance = 2005
Session: Rules; Semantic Technology; and Cross-Industry Standards
Relevance = 2004
true.

?- query([exceptions,rule-2]).
Session: RuleML-2010 Challenge
Relevance = 4022
Session: Rule-Based Distributed/Multi-Agent Systems
Relevance = 4022
Session: Rules and Inferencing
Relevance = 4020
Session: Rules and Norms
Relevance = 4017
Session: Rules and Uncertainty
Relevance = 4012
Session: Rule-based Event Processing and Reaction Rules
Relevance = 4012
Session: Rule Transformation and Extraction
Relevance = 4012
Session: General Introduction to Rules
Relevance = 4010
Session: Rules; Semantic Technology; and Cross-Industry Standards
Relevance = 4008
true.

Υποδείξεις

- Θα χρειαστεί να υλοποιηθούν αρκετά κατηγορήματα προκειμένου να υλοποιηθεί το query/1. Ενδεικτικά αναφέρονται:
 - Κατηγορήμα για το σκοράρισμα «θέματος» με βάση μία λέξη-κλειδί
 - Κατηγορήμα για το σκοράρισμα «θέματος» με βάση όλες τις λέξεις-κλειδιά
 - Κατηγορήμα για το σκοράρισμα session με βάση όλες τις λέξεις-κλειδιά και όλα τα «θέματα» και τον τίτλο του session
 - Κατηγορήμα για την «προετοιμασία» των λέξεων-κλειδιών (βλ. παραπάνω)
 - Κατηγορήμα για την εκτύπωση των αποτελεσμάτων
 - ...
- Για διευκόλυνση, ενδέχεται να χρειαστεί να χρησιμοποιήσετε τα παρακάτω ενσωματωμένα κατηγορήματα της SWI-Prolog: **tokenize_atom/2**, για μετατροπή string σε λίστα λέξεων, **sub_string(case_insensitive,String1,String2)**, για αναζήτηση του **String1** μέσα στο **String2**, χωρίς διαχωρισμό κεφαλαίων/μικρών γραμμάτων, **max_list/2**, για την εύρεση του μέγιστου λίστας, **sum_list/2**, για το άθροισμα στοιχείων λίστας, **length/2**, για την εύρεση μήκους λίστας, κλπ.

Παρατηρήσεις

- Η εργασία είναι υποχρεωτική και καλύπτει το 100% των 2.5 μονάδων του τελικού βαθμού.
 - ✓ Σε περίπτωση που δεν παραδοθεί ο φοιτητής θα έχει ως άριστα το 7,5 στις εξετάσεις και ως βάση το 4,5/7,5.
- Η εργασία μπορεί ατομικά ή σε ομάδες το πολύ 2 ατόμων.
 - ✓ Οι εργασίες θα ελεγχθούν για αντιγραφές μέσω ειδικού λογισμικού που έχει αναπτυχθεί σε γλώσσα Prolog από τον διδάσκοντα.
- Στην αρχή του αρχείου υποχρεωτικά να περιλαμβάνονται σε σχόλια ο αριθμός μητρώου και το ονοματεπώνυμο του φοιτητή ή των φοιτητών της ομάδας.
- Ο πηγαίος κώδικας του προγράμματος να γίνει upload από έναν από τους φοιτητές της ομάδας στο σύστημα **elearning.auth.gr**.
 - ✓ Το όνομα του αρχείου να είναι '**AEM.p1**', π.χ. **1234.p1**, σε περίπτωση ομάδας που αποτελείται από έναν μόνο φοιτητή, ή '**AEM1-AEM2.p1**', π.χ. **1234-678.p1**, σε περίπτωση ομάδας που αποτελείται από δύο φοιτητές.
- Ο κώδικας θα πρέπει να είναι επαρκώς σχολιασμένος.
- Ο κώδικας θα πρέπει να είναι καλά δομημένος, δηλαδή να γίνει χρήση βοηθητικών κατηγορημάτων όπου κρίνεται απαραίτητο.
 - ✓ Θα βαθμολογηθεί και το στυλ προγραμματισμού, όχι μόνο το σωστό αποτέλεσμα.
 - ✓ Οποιοσδήποτε βελτιώσεις ή προσθήκες στις απαιτήσεις της εργασίας πραγματοποιηθούν, να επισημανθούν με σχόλια στην αρχή του κώδικα, ώστε να αξιολογηθούν επιπλέον.