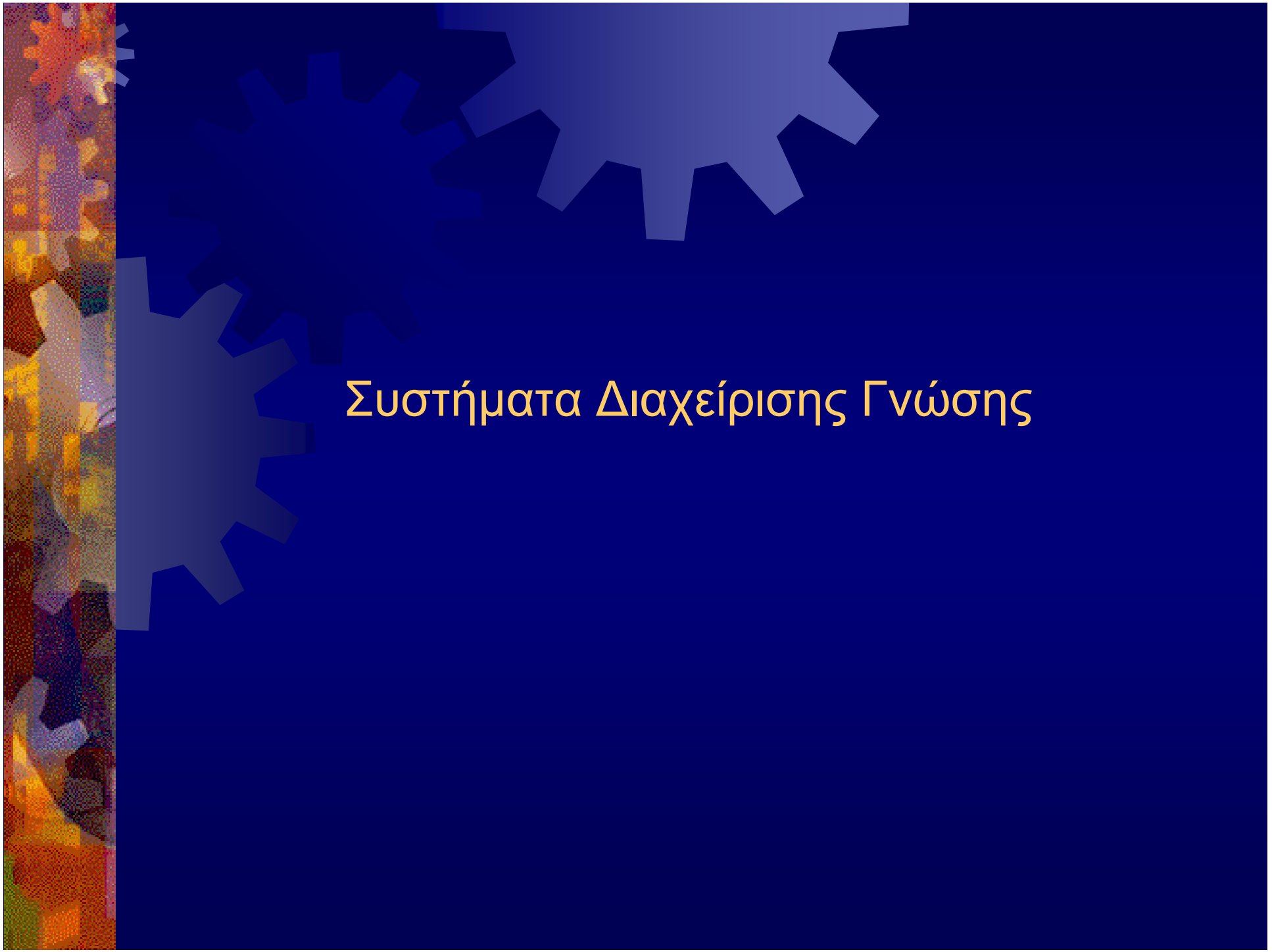

Συστήματα Διαχείρισης Γνώσης

Εισαγωγή

- Νέα κατηγορία πληροφοριακών συστημάτων που έχουν τη δυνατότητα να υποστηρίζουν τους χρήστες τους στην επίλυση προβλημάτων και λήψη αποφάσεων
  - δεν τους υποκαθιστούν
  - προκύπτουν από τη σύγκλιση των έμπειρων συστημάτων και των συστημάτων υποστήριξης ομάδων
- Διάφορες ονομασίες
  - Σ.Υ.Α. Βασιζόμενα στη Γνώση (Knowledge-Based Decision Support Systems - KB-DSS) (Klein and Methlie, 1990; Guida et al., 1992; Liberatore and Stylianou, 1993)
  - Ευφυή Συστήματα Υποστήριξης Αποφάσεων (Intelligent Decision Support Systems - IDSS) (McGovern et al., 1991; Gottinger and Weimann, 1992)
  - Βασιζόμενα στη Γνώση Συστήματα Υποστήριξης του Μάνατζμεντ (Knowledge Based Management Support Systems) (Doukidis et al., 1989)
  - Έμπειρα Συστήματα Υποστήριξης (Expert Support Systems - ESS) (Van Weelderen and Sol, 1993).
- Βασικό χαρακτηριστικό:
  - συστήματα που παρέχουν επεξεργασμένη πληροφορία, η οποία έχει αποκτηθεί από πολλές πηγές (ειδικούς, βάσεις δεδομένων, file servers, κλπ.)
  - δεν επιδιώκουν να μοντελοποιήσουν το τρόπο λήψης απόφασης

Η γνώση στο επιχειρησιακό περιβάλλον

- Ορισμός (λεξικό Webster) :

« Γνώση είναι το γεγονός ή συνθήκη της αντίληψης που έχει επέλθει μέσα από εμπειρίες ή συσχέτιση »

- Έμφαση...

στην αποτελεσματική χρήση και στο αποτέλεσμα (όχι στην ανακάλυψη της αλήθειας)

Κατηγορίες γνώσης (Nonaka and Takeuchi, 1995)

Ρητή γνώση

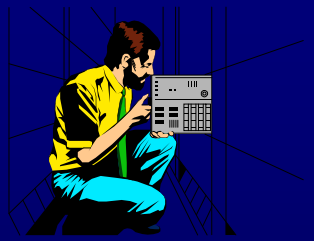
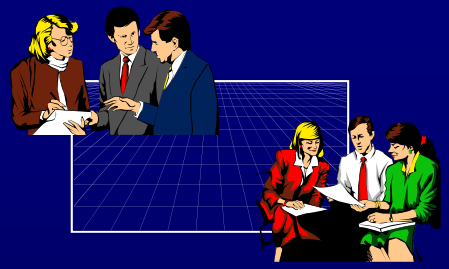
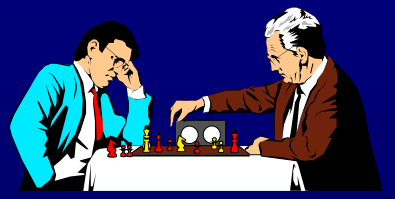
- Δομημένη και αντικειμενική
- Αρθρωμένη σε δομημένη γλώσσα
- Μπορεί να αποτυπωθεί σε κείμενα, διαδικασίες ή βάσεις δεδομένων
- Γλώσσα – Πληροφορία – Μέσο μεταφοράς

Άρρητη γνώση

- Αδόμητη και υποκειμενική
- Ενσωματωμένη σε προσωπική εμπειρία
- Ένστικτο – Διόραση - Ικανότητες

Γνώση σε οργανισμούς

- Σε οργανισμούς, συχνά ενσωματώνεται όχι μόνο σε κείμενα ή βάσεις δεδομένων αλλά και σε επιχειρησιακές λειτουργίες, πρακτικές και κανόνες και ομάδες εργαζομένων

Άτομα | Συστήματα | Δίκτυα | Λειτουργίες

Η ανάγκη για διαχείριση γνώσης

Αυξημένος ανταγωνισμός

Ανάγκη για συνεχή ανανέωση και καινοτομία

Προϊόντα ενσωματώνουν γνώση για διαφοροποίηση

Παγκοσμιοποίηση

Εξαγορές και συγχωνεύσεις

Γεωγραφική διασπορά

Ευκολότερη επικοινωνία και διασύνδεση

Ανάγκη για διαχείριση γνώσης

Διαρθρωτικές Αλλαγές

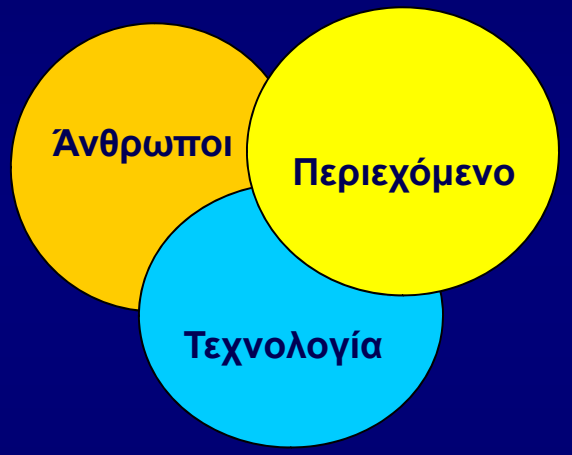
Έχουν οδηγήσει σε γνωστικά κενά

Αυξημένη κινητικότητα υπαλλήλων

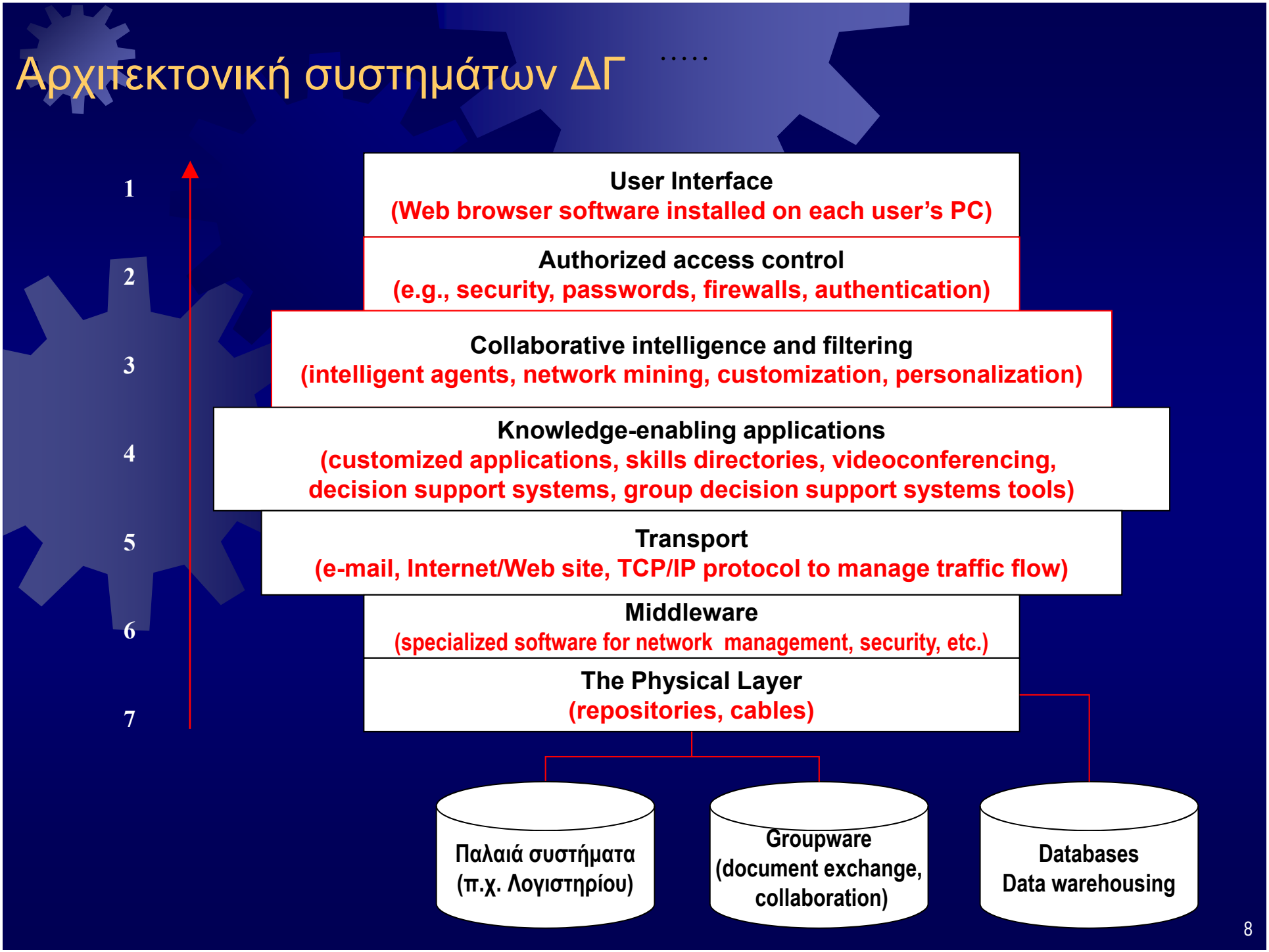
Γνώση απαιτεί χρόνο και εμπειρίες για να αποκτηθεί

Παράγοντες επιρροής ΔΓ

- Ανθρώπινος πυρήνας: Οργάνωση, κουλτούρα, διαδικασίες, κλπ.
- Πυρήνας περιεχομένου: Προσδιορισμός και καταγραφή του γνωστικού ενεργητικού (knowledge assets)
- Τεχνολογικός πυρήνας: Η συνολική τεχνολογική υποδομή που απαιτείται για την λειτουργία της ΔΓ

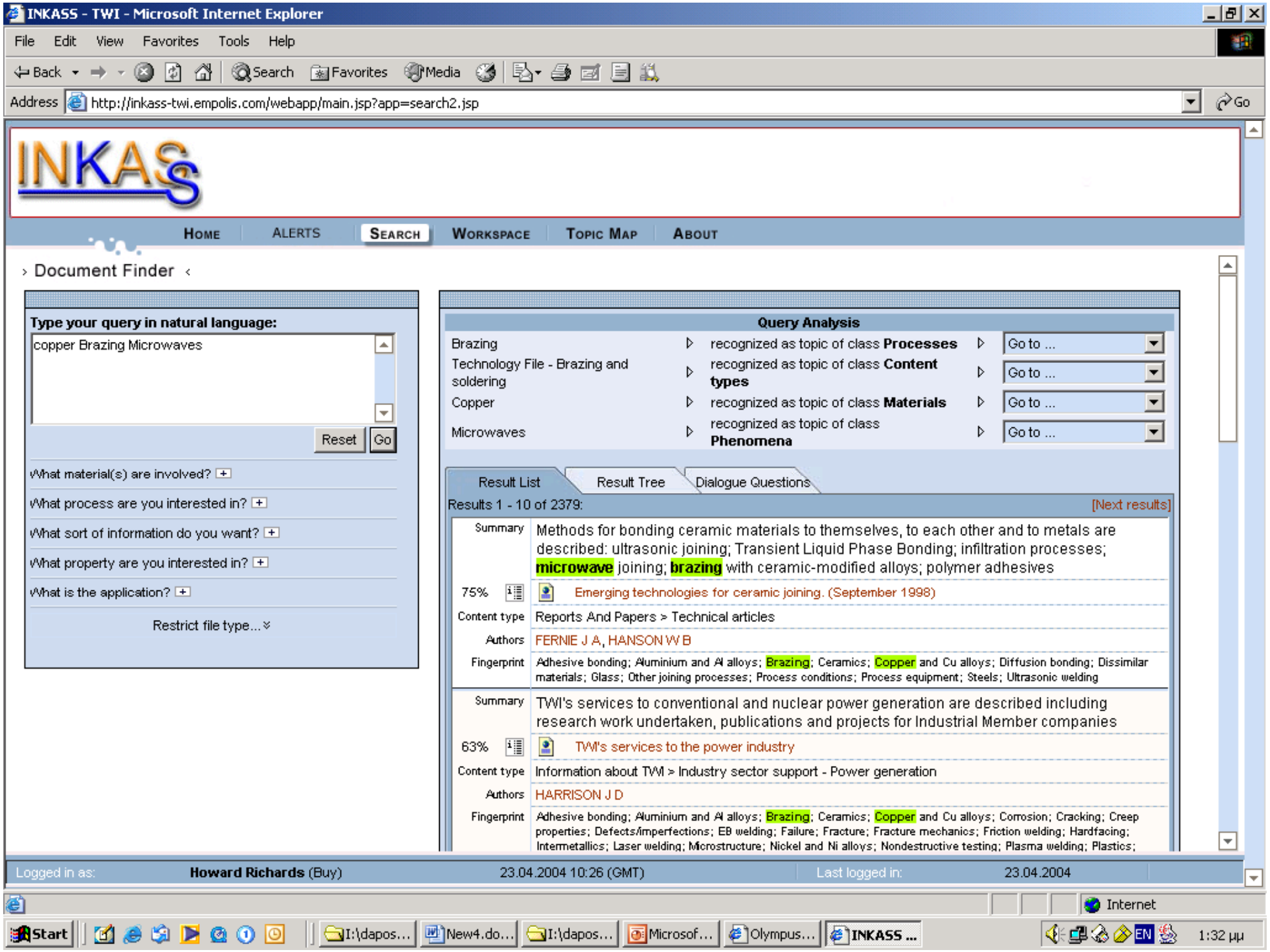


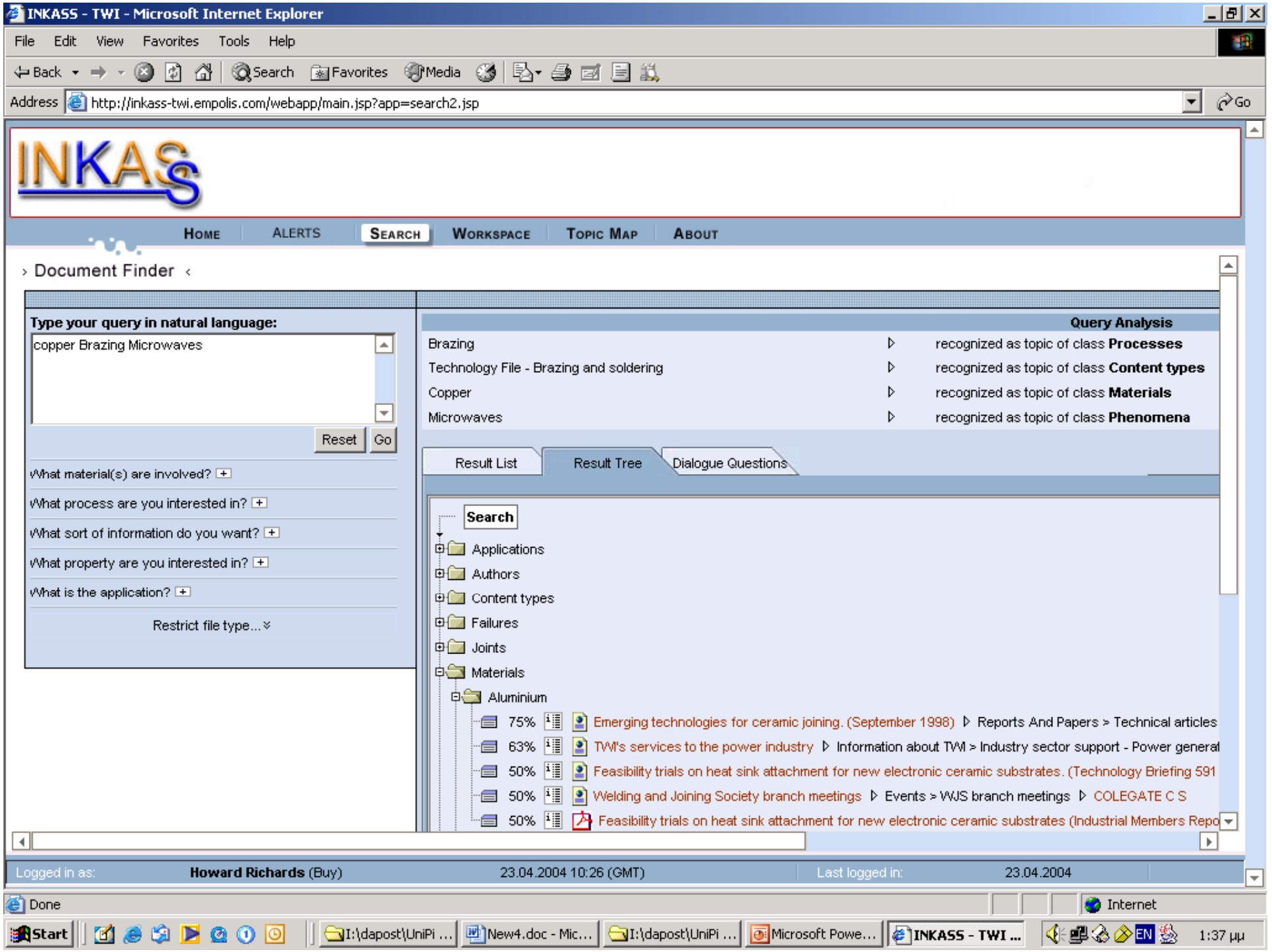
Αρχιτεκτονική συστημάτων ΔΓ .....

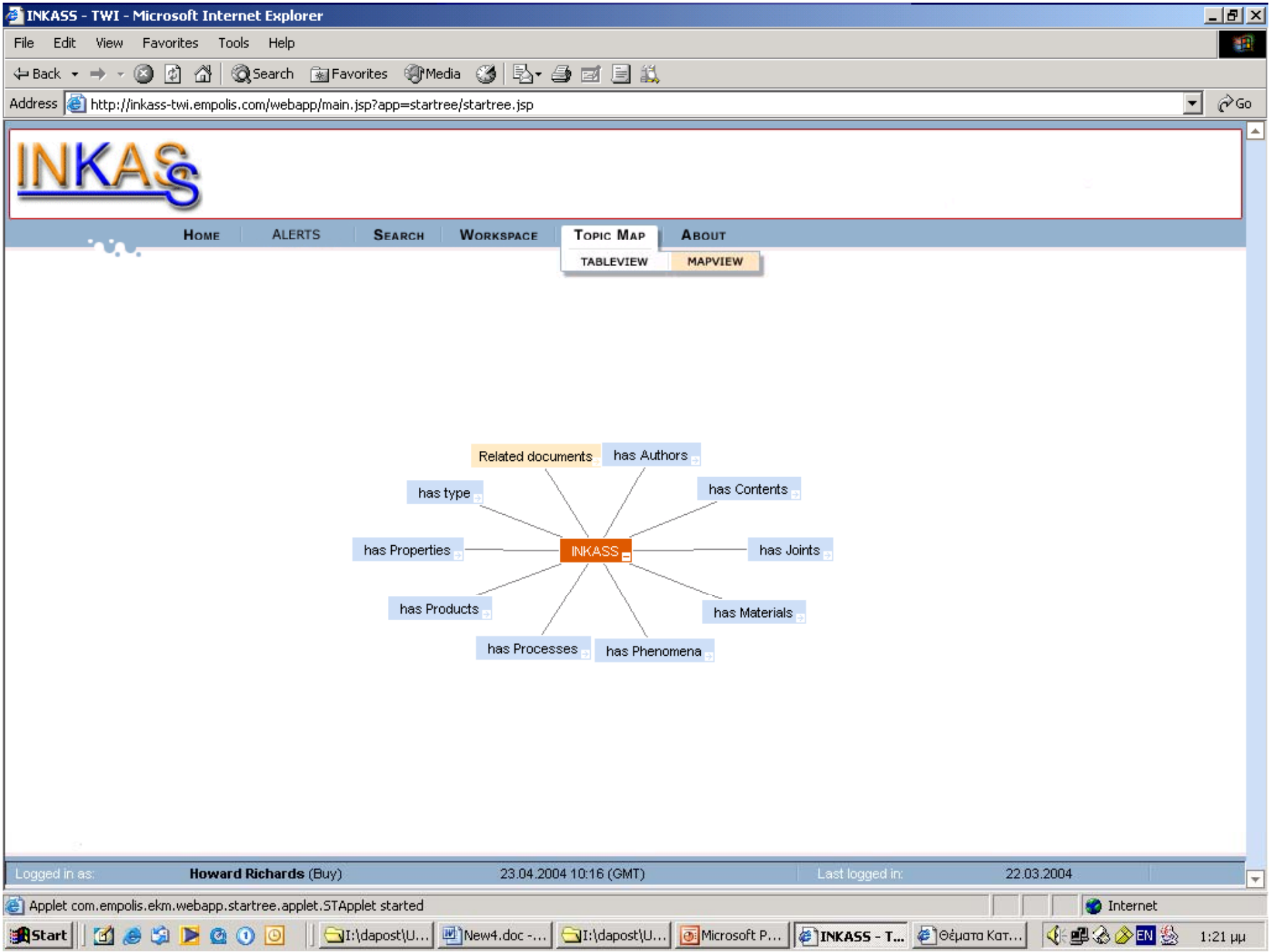


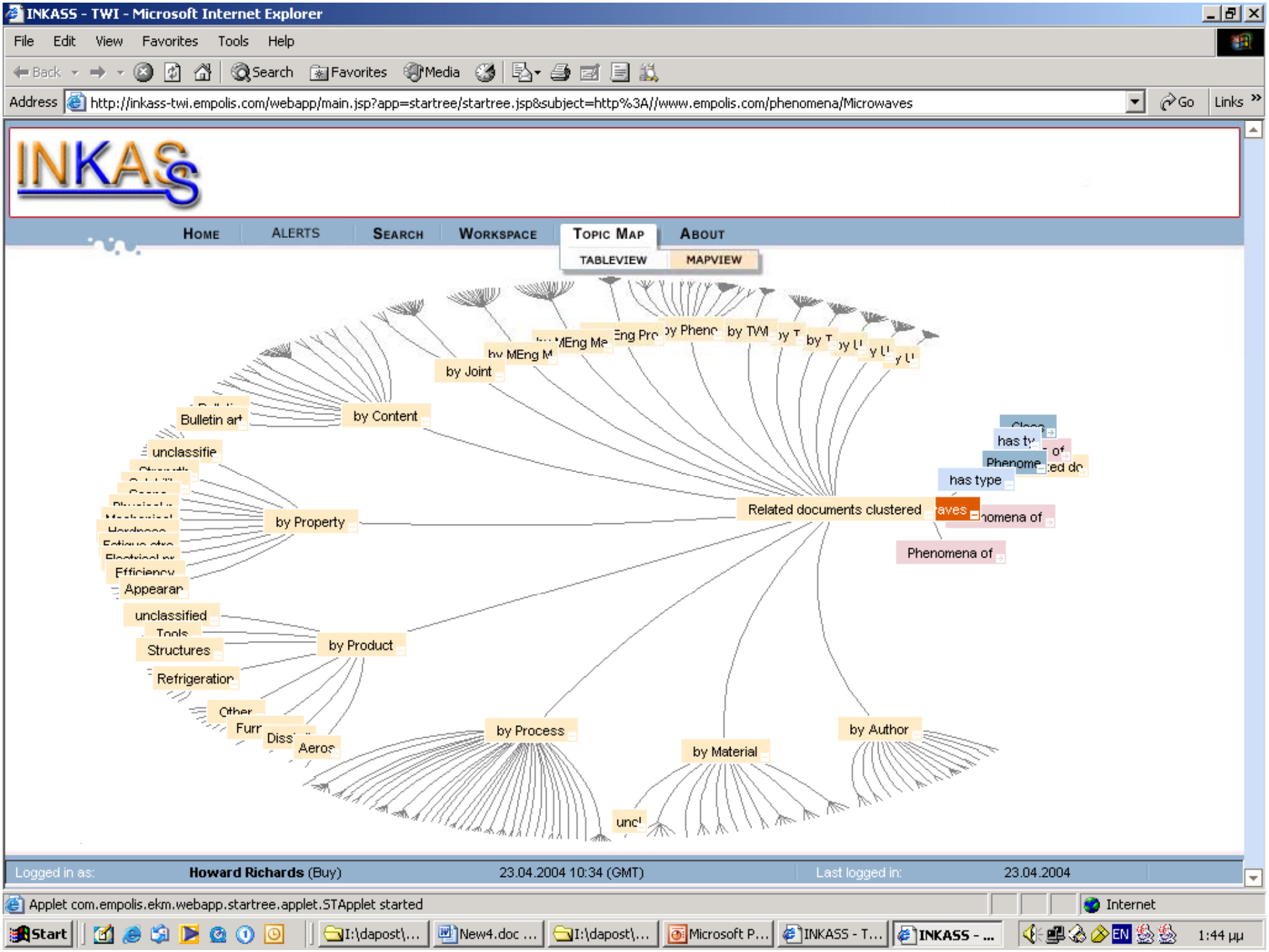
Επίπεδο διεπαφής χρηστών

- Έμφαση και στην άρρητη γνώση
  - Links σε ανθρώπους, email, σύγχρονη/ασύγχρονη επικοινωνία, video
- Εστίαση σε σχεδιασμό user interface με
  - Συνέπεια
    - Σημασία μενού, εικονιδίων, κουμπιών
    - Μορφή πληροφορίας
  - Συνάφεια
    - Σχετικές πληροφορίες
    - Προσαρμοστικότητα
    - Εξατομίκευση
  - Οπτική σαφήνεια
    - Ολόκληρες οθόνες, δεσμοί, κλπ.
  - Πλοήγηση
  - Χρηστικότητα


INKASS - TWI - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://inkass-twi.empolis.com/webapp/main.jsp?app=search2.jsp
Go
INKASS
HOME ALERTS SEARCH WORKSPACE TOPIC MAP ABOUT
› Document Finder ‹
Type your query in natural language:
copper Brazing Microwaves
Reset Go
What material(s) are involved?
What process are you interested in?
What sort of information do you want?
What property are you interested in?
What is the application?
Restrict file type...
Query Analysis
Brazing recognized as topic of class Processes Go to ...
Technology File - Brazing and soldering recognized as topic of class Content types Go to ...
Copper recognized as topic of class Materials Go to ...
Microwaves recognized as topic of class Phenomena Go to ...
Result List Result Tree Dialogue Questions
Results 1 - 10 of 2379: [Next results]
Summary Methods for bonding ceramic materials to themselves, to each other and to metals are described: ultrasonic joining; Transient Liquid Phase Bonding; infiltration processes; microwave joining; brazing with ceramic-modified alloys; polymer adhesives
75% Emerging technologies for ceramic joining. (September 1998)
Content type Reports And Papers > Technical articles
Authors FERNIE J A, HANSON W B
Fingerprint Adhesive bonding; Aluminium and Al alloys; Brazing; Ceramics; Copper and Cu alloys; Diffusion bonding; Dissimilar materials; Glass; Other joining processes; Process conditions; Process equipment; Steels; Ultrasonic welding
Summary TWI's services to conventional and nuclear power generation are described including research work undertaken, publications and projects for Industrial Member companies
63% TWI's services to the power industry
Content type Information about TWI > Industry sector support - Power generation
Authors HARRISON J D
Fingerprint Adhesive bonding; Aluminium and Al alloys; Brazing; Ceramics; Copper and Cu alloys; Corrosion; Cracking; Creep properties; Defects/imperfections; EB welding; Failure; Fracture; Fracture mechanics; Friction welding; Hardfacing; Intermetallics; Laser welding; Microstructure; Nickel and Ni alloys; Nondestructive testing; Plasma welding; Plastics;
Logged in as: Howard Richards (Buy) 23.04.2004 10:26 (GMT) Last logged in: 23.04.2004
Internet
Start I:\dapos... New4.do... I:\dapos... Microsof... Olympus... INKASS ... 1:32 μμ


INKASS - TWI - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://inkass-twi.empolis.com/webapp/main.jsp?app=search2.jsp
Go
INKASS
HOME ALERTS SEARCH WORKSPACE TOPIC MAP ABOUT
› Document Finder ‹
Type your query in natural language:
copper Brazing Microwaves
Reset Go
What material(s) are involved?
What process are you interested in?
What sort of information do you want?
What property are you interested in?
What is the application?
Restrict file type...
Query Analysis
Brazing recognized as topic of class Processes
Technology File - Brazing and soldering recognized as topic of class Content types
Copper recognized as topic of class Materials
Microwaves recognized as topic of class Phenomena
Result List
Result Tree
Dialogue Questions
Search
Applications
Authors
Content types
Failures
Joints
Materials
Aluminium
75% Emerging technologies for ceramic joining. (September 1998) ▷ Reports And Papers > Technical articles
63% TWI's services to the power industry ▷ Information about TWI > Industry sector support - Power general
50% Feasibility trials on heat sink attachment for new electronic ceramic substrates. (Technology Briefing 591
50% Welding and Joining Society branch meetings ▷ Events > WJS branch meetings ▷ COLEGATE C S
50% Feasibility trials on heat sink attachment for new electronic ceramic substrates (Industrial Members Repo
Logged in as: Howard Richards (Buy) 23.04.2004 10:26 (GMT) Last logged in: 23.04.2004
Done
Internet
Start
I:\dapost\UniPi ...
New4.doc - Mic...
I:\dapost\UniPi ...
Microsoft Powe...
INKASS - TWI ...
1:37 μμ


INKASS - TWI - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://inkass-twi.empolis.com/webapp/main.jsp?app=startree/startree.jsp
Go
INKASS
HOME
ALERTS
SEARCH
WORKSPACE
TOPIC MAP
ABOUT
TABLEVIEW
MAPVIEW
Related documents
has Authors
has type
has Contents
has Properties
INKASS
has Joints
has Products
has Materials
has Processes
has Phenomena
Logged in as: Howard Richards (Buy)
23.04.2004 10:16 (GMT)
Last logged in: 22.03.2004
Applet com.empolis.ekm.webapp.startree.applet.STApplet started
Internet
Start
I:\dapost\U...
New4.doc -...
I:\dapost\U...
Microsoft P...
INKASS - T...
Θέματα Κατ...
EN
1:21 μμ


INKASS - TWI - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://inkass-twi.empolis.com/webapp/main.jsp?app=startree/startree.jsp&subject=http%3A//www.empolis.com/phenomena/Microwaves
Go Links
INKASS
HOME ALERTS SEARCH WORKSPACE TOPIC MAP ABOUT
TABLEVIEW MAPVIEW
by Joint
by Content
Bulletin art
unclassifie
by Property
Appearar
unclassified
Structures
by Product
Refrigeration
by Process
by Material
by Author
Related documents clustered
has type
Phenomena of
Logged in as: Howard Richards (Buy) 23.04.2004 10:34 (GMT) Last logged in: 23.04.2004
Applet com.empolis.ekm.webapp.startree.applet.STApplet started
Internet
Start
1:44 μμ

Επίπεδο ελέγχου πρόσβασης

- Εξασφαλίζει εξουσιοδοτημένη πρόσβαση στο σύστημα, στη γνώση, στα αρχεία
- Η πρόσβαση μπορεί να γίνει (συνήθως) στα εξής επίπεδα
  - Intranet
    - Εσωτερικό δίκτυο που αναπτύσσεται και βασίζεται σε τεχνολογίες Internet
    - Κοινό, οικονομικό και συμβατό λογισμικό
  - Extranet
    - Intranets με τις απαραίτητες προεκτάσεις που επιτρέπουν σε συγκεκριμένους πελάτες ή συνεργάτες να έχουν πρόσβαση σε εταιρικά δεδομένα
  - Internet

Χαρακτηριστικά/περιορισμοί firewalls

- Ειδικό λογισμικό που προστατεύει το intranet από εξωτερικές επιθέσεις
- Προστασία έναντι:
  - Ανεπιθύμητα e-mails
  - Μη εξουσιοδοτημένη πρόσβαση
    - Πρόσβαση από IPs εκτός domain οργανισμού
  - Είσοδο/έξοδο ανεπιθύμητου υλικού
- Περιορισμοί firewalls:
  - Προσβολές που παρακάμπτουν firewall
    - Αδύναμες πολιτικές ασφαλείας
    - floppy disks/USBs
    - Ανθρώπινος παράγοντας

Επίπεδο συνεργατικής νοημοσύνης και φιλτραρίσματος (Collaborative Intelligence & Filtering Layer)

- Παρέχει προσωποποιημένες όψεις στην αποθηκευμένη πληροφορία
- Μειώνει χρόνο αναζήτησης
- Ενεργή βοήθεια στην αναζήτηση και πρόσβαση σε πληροφορίες μέσω agents
- Παροχή «ευφυούς» βοήθειας σε εξειδικευμένες εφαρμογές:
  - Οργάνωση συναντήσεων
  - Υποδείξεις, προτάσεις, υπενθυμίσεις
  - ...

Amazon.com: Books: Knowledge Asset Management - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://www.amazon.com/exec/obidos/tg/detail/-/1852335831/qid=1082721217/sr=1-1/ref=sr_1_1/103-4462710-0333415?v=glance&s=books Go

(Beta-What is this?)

WELCOME | YOUR STORE | BOOKS | APPAREL & ACCESSORIES | ELECTRONICS | TOYS & GAMES | MUSIC | CELL PHONES & SERVICE | SEE MORE STORES | Your Gold Box

SEARCH | BROWSE SUBJECTS | BESTSELLERS | MAGAZINES | CORPORATE ACCOUNTS | E-BOOKS & DOCS | BARGAIN BOOKS | USED BOOKS

Search Inside the Book™

SEARCH

Books GO!

WEB SEARCH

GO!

Powered by Google

BOOK INFORMATION

buying info

editorial reviews

RATE THIS BOOK

(Sign in to rate this item.)

Favorite Magazines!

Knowledge Asset Management

by Gregoris N. Mentzas, Dimitris Apostolou (Editor), Andreas Abecker (Editor), Ron Young (Editor)

List Price: $99.00

Price: **$99.00** & This item ships for **FREE with Super Saver Shipping**. See details.

Availability: This title usually ships within 1 to 3 weeks. Please note that special order titles occasionally go out of print, or publishers run out of stock. These hard-to-find titles are not discounted and are subject to an additional charge of $1.99 per book due to the extra cost of ordering them. We will notify you within 2-3 weeks if we have trouble obtaining this title

9 used & new from $65.00

Edition: Hardcover

See more product details

READY TO BUY?

Add to Shopping Cart

or

Sign in to turn on 1-Click ordering.

MORE BUYING CHOICES

9 used & new from $65.00

Have one to sell? Sell yours here

Add to Wish List

Add to Wedding Registry

Don't have one? We'll set one up for you.

Customers interested in Knowledge Asset Management may also be interested in:

Sponsored Links (What's this?) Feedback

- **Knowledge Management** Kit
  Blueprints, templates, plans. Full thorough toolkit for immediate use
  ovitztaylorgates.com

- **Knowledge Management**
  Mobile **knowledge management** software. Free Trial
  www.ptshome.com/tracerpluspro.htm

http://www.amazon.com/exec/obidos/search-handle-url/index=books&field-author=Gregoris N. Mentzas/103-4462710-0333415 Internet

Start | I:\dapost\U... | New4.doc -... | I:\dapost\U... | Microsoft P... | Amazon.c... | EN | 2:58 μμ

Ευφυή συστήματα

- Προγράμματα που:
  - Προσομοιάζουν λογική ειδικών σε συγκεκριμένους τομείς
  - Κωδικοποιούν και χειρίζονται τη γνώση και τη συλλογιστική ενός ειδικού σε σκοπό την επίλυση προβλημάτων
- Χρησιμοποιούνται ουσιαστικά με δύο τρόπους
  - Για παροχή γνώσης σε μη ειδικούς, απουσία των ειδικών
  - Συμβουλευτικά σε ειδικούς, υποβοηθώντας τους
- Βασικά στοιχεία:
  - Justifier
    - Εξηγεί πως και γιατί δίδεται μία απάντηση
  - Inference engine
    - μηχανισμός αντιμετώπισης προβλημάτων για επίλυση και εξαγωγή συμπερασμάτων
  - Scheduler
    - εναρμονίζει και ελέγχει επεξεργασία κανόνων (rule processing)

Επίπεδο εφαρμογών υποστήριξης της γνώσης

- Παρέχει εφαρμογές που στοχεύουν άμεσα στη ΔΓ και μάθηση
- Παρέχει πρόσβαση σε:
  - βάσεις γνώσης
  - εφαρμογές μάθησης
  - παροχή γνώσης στο επίπεδο επιχειρησιακών λειτουργιών
  - παροχή γνώσης στο επίπεδο έργων
  - εφαρμογές προσωπικής ανάπτυξης
  - career planning

Επίπεδο μεταφοράς (Transport Layer)

- Τεχνικό επίπεδο
- Περιλαμβάνει LANs, WANs, intranets, extranets, and the Internet
- Εξετάζει τεχνική δυνατότητα υποστήριξης multimedia, graphics
- Βασικά κριτήρια: Εύρος ζώνης και ταχύτητες σύνδεσης

Ενδιάμεσο επίπεδο (Middleware Layer)

- Λογισμικό που παρέχει σύνδεση όλων των εφαρμογών και λογισμικού
- Υποστηρίζει σύνδεση και interfacing με παλαιά συστήματα και προγράμματα που «τρέχουν» σε άλλες πλατφόρμες
- Εστίαση σε παλαιές εφαρμογές που τροφοδοτούν σύστημα ΔΓ

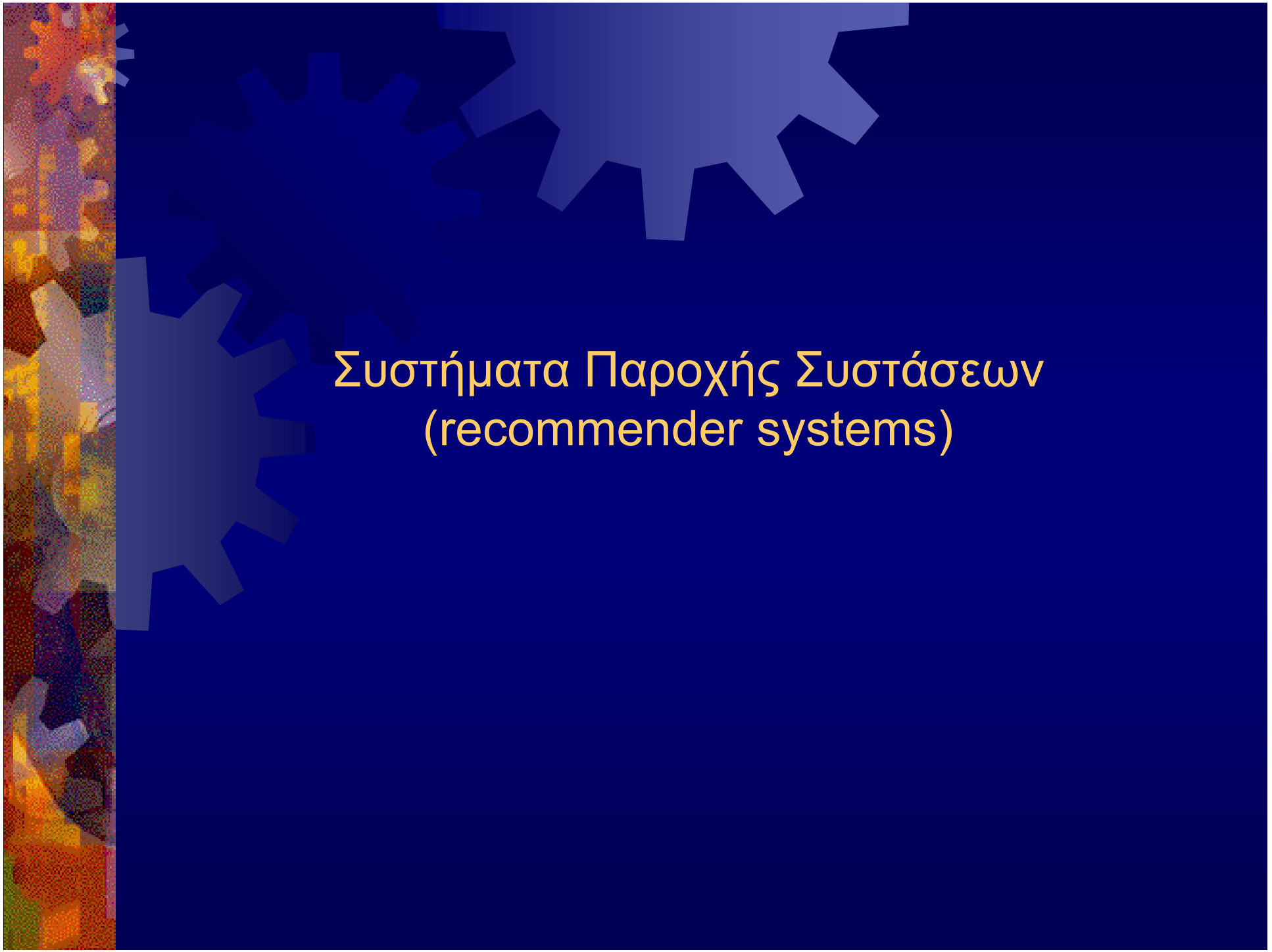
Επίπεδο αποθήκευσης (Repositories Layer)

- Χαμηλότερο επίπεδο αρχιτεκτονικής ΔΓ
- Αντιπροσωπεύει το φυσικό επίπεδο στο οποίο τοποθετούνται τα repositories
- Περιλαμβάνει:
  - data warehouses
  - legacy applications
  - operational databases
  - ειδικές εφαρμογές για διαχείριση ασφάλειας και διακίνησης/φορτίου

Ανάπτυξη ή αγορά;

- Μέχρι τέλη 90, σοβαρές εφαρμογές ΔΓ ήταν custom-built
- Σήμερα ώριμη αγορά
- Τάση για ευκολόχρηστα και εύκολα στην εφαρμογή, γενικευμένα συστήματα
- Εξατομίκευση (customisation) ή ανάπτυξη συνήθως για εφαρμογές Knowledge-Enabling Application Layer
- Πρέπει να τίθενται κριτήρια επιλογής
- Σημαντικό θέμα ποιος έχει ευθύνη επιλογής και (κυρίως) διαχείρισης και συντήρησης)

Δυνατότητα	*Κόστος*	*Χρόνος*	*Προσαρμογή*	*Σχόλια*
Εσωτερική ανάπτυξη	Συνήθως υψηλό	Περισσότερο χρονοβόρα από την εγκατάσταση έτοιμου πακέτου	Υψηλή, εξαρτάται από την εμπειρία του προσωπικού	Μπορεί να γίνει πρόσληψη συμβούλων για τη βελτίωση της ποιότητας αλλά αυξάνεται το κόστος. Ενδεχομένως να υπάρξει διαρροή κρίσιμων πληροφοριών από τους συμβούλους.
Ανάθεση σε εξωτερικό φορέα	Μεσαίο προς υψηλό	Συντομότερα από την εσωτερική ανάπτυξη	Υψηλή	Ο ανταγωνιστής μπορεί ήδη να έχει το ίδιο σύστημα ΔΓ.
Αγορά έτοιμου πακέτου	Χαμηλό προς μεσαίο	Μηδενικός	Συνήθως είναι χρησιμοποιήσιμο σε ποσοστό έως 80%	Η εγκατάσταση γίνεται σχετικά εύκολα και γρήγορα.


Συστήματα Παροχής Συστάσεων
(recommender systems)

Συστήματα Παροχής Συστάσεων (recommender systems)

1. Ο χρήστης εκφράζει την γνώμη του (opinion) για ένα ή περισσότερα αντικείμενα

- Ανάλογα με το πώς εκφράζεται η γνώμη:
  - **Ρητά (Explicit)**: Ο χρήστης βαθμολογεί (rates) τα αντικείμενα
  - **Έμμεσα (Implicit)**: Με βάση την χρήση (usage) του συστήματος

2. Το recommender σύστημα προτείνει στον χρήστη άλλα παρόμοια (similar) αντικείμενα

- Ανάλογα με το πώς ορίζεται το «παρόμοια αντικείμενα»:
  - **Content based filtering**: Παρόμοια σε περιεχόμενο
  - **Collaborative filtering**: Παρόμοια ως προς την εκτίμηση που έχουν από άλλους χρήστες
  - **Υβριδικά:** Συνδυασμός των παραπάνω

Κατηγορίες Συστημάτων Παροχής Συστάσεων

Ακόμα μια κατηγοριοποίηση:

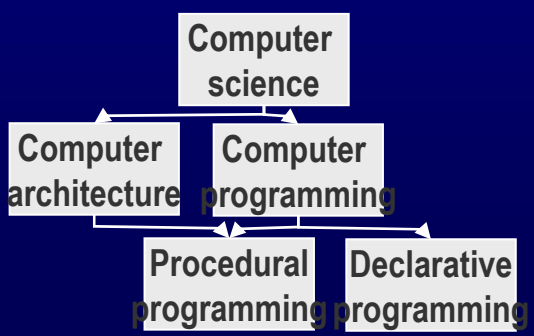
- Memory-Based
  - Το σύστημα χρησιμοποιεί heuristics και υπολογίζει επιτόπου τις προτάσεις του με βάση την δραστηριότητα του χρήστη
- Model-Based
  - Το σύστημα χρησιμοποιεί στατιστικές και machine learning μεθόδους για να δημιουργήσει ένα predictive model, το οποίο τελικά το χρησιμοποιεί για να προτείνει.

Explicit vs Implicit opinion

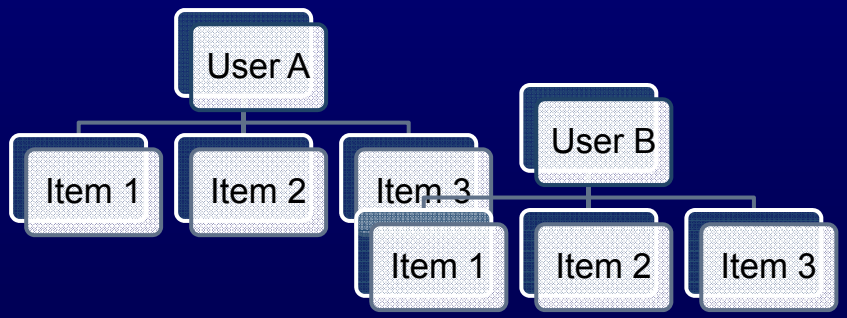
- Explicit opinion:
  - Ακρίβεια (π.χ. 3/5 stars)
  - Θετικό και αρνητικό feedback
  - Δύσκολα συλλέγεται (ο χρήστης πρέπει να δώσει μόνος του feedback)
- Implicit opinion:
  - Όχι τόση ακρίβεια (noisy)
  - Συνήθως μόνο θετικό feedback
  - Συλλέγεται εύκολα (αυτόματα)
- Τρόποι αναγνώρισης implicit opinion ανάλογα με το πώς ένας χρήστης χρησιμοποιεί το σύστημα
  - Κάνει click σε συγκεκριμένες σελίδες και με συγκεκριμένη σειρά, ενώ αγνοεί κάποιες άλλες
  - Διαβάζει κάποια αρχεία περισσότερο
  - Αποθηκεύει ή τυπώνει κάποια αρχεία
  - Κατεβάζει αρχεία περιεχομένου ή τα προσθέτει σε bookmarks

Content-based vs Collaborative Filtering

- Content-based filtering
  - Αναλύουν το περιεχόμενο αρχείου
  - Οι χρήστες και τα αντικείμενα περιεχομένου αντιπροσωπεύονται από ένα σύνολο χαρακτηριστικών
- Collaborative filtering
  - **ΔΕΝ** χρειάζονται πληροφορία για το περιεχόμενο των αντικειμένων περιεχομένου
  - Βασίζονται στην υπόθεση ότι ένας χρήστης ενδιαφέρεται για αντικείμενα περιεχομένου που προτιμούν παρόμοιοι χρήστες



Content-based filtering



collaborative filtering

Επεξήγηση των recommendations

- Και οι δύο κατηγορίες συστημάτων είναι καλό να επεξηγούν για ποιον λόγο έκαναν recommend ένα item
  - This is a current hit …
  - More on this author …
  - Try something from similar authors …
  - Someone similar to you also like this …
  - These two go together …
  - This is most popular in your group …
  - This is highly rated …
  - Try something new …

Content based Filtering Systems (1/3)

- **Τι κάνουν**: Προτείνουν items με βάση το τι άρεσε στον χρήστη στον παρελθόν
  - Βασική ιδέα: ένας χρήστης είναι πιθανό να έχει την ίδια γνώμη για similar items
- **Πως?**

1. Το σύστημα μαθαίνει τις προτιμήσεις του χρήστη μέσω του user feedback (Explicit ή Implicit) και χτίζει ένα user profile…
   - Το profile περιλαμβάνει πληροφορία για τα items of interest του χρήστη, π.χ. συγκεκριμένα άρθρα, βιβλία, κτλ.
2. …Και βρίσκει similar items τα οποία και προτείνει στον χρήστη
   - Το input είναι τα items of interest που βρίσκονται στο user profile και το output τα recommendations…

Content based Filtering Systems (2/3)

- Η όλη διαδικασία μπορεί να θεωρηθεί σαν μια αναζήτηση εγγράφων στις μηχανές αναζήτησης
  - Το user profile με τα items of interest είναι το query
  - Τα items που περιέχονται στο KMS είναι το document base από το οποίο ψάχνουμε να βρούμε τα similar items που θα γίνουν recommend

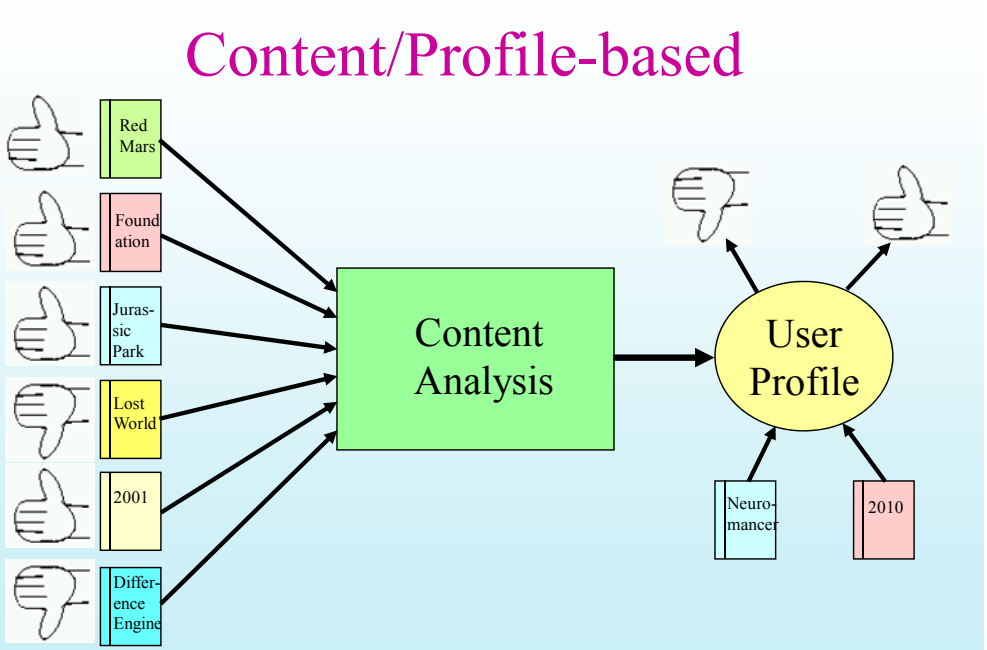
- Πως αναγνωρίζεται ένα item σαν similar item?

1. Τα items περιγράφονται με βάση:
   - Τα χαρακτηριστικά τους
     - π.χ. Ένα έγγραφο έχει α) γνωστική περιοχή, β) συγγραφέα, γ) λέξεις-κλειδιά κτλ
   - Κάποια άλλα free tags που τα περιγράφουν
   - Το κείμενο που περιγράφει το item

Content based Filtering Systems (3/3)

2. Τα items που έχουν παρόμοια χαρακτηριστικά ή tags ή περιγράφονται με παρόμοιο κείμενο, θεωρούνται similar
   - Δεν έχουν όλες οι λέξεις (χαρακτηριστικά, tags ή keywords του κειμένου) που περιγράφουν ένα item την ίδια βαρύτητα
   - Πιο σημαντικές θεωρούνται οι λέξεις
     - Που εμφανίζονται πιο συχνά από τον μέσο όρο
     - Που είναι πιο περιγραφικές για κάποια κλάση αντικειμένων
   - Π.χ. για περιγραφές restaurant, λέξεις όπως “noodle”, “shrimp”, “basil”, “exotic”, “salmon” είναι σημαντικές

- Για να συμπεράνει το σύστημα ποια items είναι similar χρησιμοποιούνται τεχνικές ανάλυσης περιεχομένου και machine learning
  - Tf–idf weight (term frequency–inverse document frequency)



Παράδειγμα Tf–idf

- Ο χρήστης έχει εκφράσει την γνώμη του για τα Documents 1-4
- Οι περιγραφές των documents περιέχουν κάποια keywords
- Πρόβλημα: Ποιο προβλέπεται να είναι το rating του χρήστη για το document 5?

Document / Keyword	hardware	processor	cpu	linux	windows	**Rating**
Document 1	x2	x2	x3	x2	x1	**3**
Document 2		x1	x1			**5**
Document 3	x1		x1			**8**
Document 4	x3	x1		x2		**1**
Document 5		x2	x1		x1	**?**

Παράδειγμα Tf–idf

- The **term count** in the given document is simply the number of times a given term appears in that document.
  - This count is usually normalized to prevent a bias towards longer documents (which may have a higher term count regardless of the actual importance of that term in the document) to give a measure of the importance of the term ti within the particular document dj.

$$\mathrm{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

- where ni,j is the number of occurrences of the considered term (ti) in document dj, and the denominator is the sum of number of occurrences of all terms in document dj, that is, the size of the document | dj | .

Παράδειγμα Tf–idf

- The inverse document frequency is a measure of the general importance of the term (obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient).

$$\mathrm{idf}_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|}$$

- With
  - | D | : cardinality of D, or the total number of documents in the corpus
  - $|\{d \in D : t \in d\}|$: number of documents where the term ti appears.
  - If the term is not in the corpus, this will lead to a division-by-zero. It is therefore common to adjust the denominator to: $1 + |\{j : t_i \in d_j\}|$

Παράδειγμα Tf–idf

- Για παράδειγμα στο Document 1, η λέξη hardware εμφανίζεται 2 φορές ενώ συνολικά το Document 1 περιέχει 10 λέξεις.
  - Άρα TF = 2/10 = 0.2
- Η λέξη hardware εμφανίζεται σε 3 από τα συνολικά 5 documents
  - Άρα IDF = log (5/3) = 0.22
  - Tf-idf = TF*IDF για τη λέξη hardware = 0.2 * 0.22 = 0.044
- Ομοίως για τις υπόλοιπες λέξεις στο Document 1

Document / Keyword	hardware	processor	cpu	linux	windows	**Rating**
Document 1	x2	x2	x3	x2	x1	**3**
Document 2		x1	x1			**5**
Document 3	x1		x1			**8**
Document 4	x3	x1		x2		**1**
Document 5		x2	x1		x1	**?**

Παράδειγμα Tf–idf

- Στη συνέχεια αναπαριστούμε όλα τα Documents σαν διανύσματα βαρών:
  - $V_{Document\ 1}$ = {0.044, …, …., ….}
  - $V_{Document\ 2}$ = {…., …, …., ….}
  - …
  - $V_{Document\ 5}$ = {…., …, …., ….}

$$\mathbf{v}_d = [w_{1,d}, w_{2,d}, \ldots, w_{N,d}]^T$$

- όπου βάρη ($w_{i,d}$) είναι οι τιμές Tf-idf που υπολογίσαμε στο προηγούμενο βήμα.
- Υπολογίζουμε την ομοιότητα μεταξύ documents dj και q με τον εξής τρόπο:

$$sim(d_j, q) = \frac{\mathbf{d_j} \cdot \mathbf{q}}{\|\mathbf{d_j}\| \, \|\mathbf{q}\|} = \frac{\sum_{i=1}^{N} w_{i,j} * w_{i,q}}{\sqrt{\sum_{i=1}^{N} w_{i,j}^2} * \sqrt{\sum_{i=1}^{N} w_{i,q}^2}}$$

Παράδειγμα Tf–idf

- Τέλος υπολογίζουμε το εκτιμούμενο rating που θα έδινε ο χρήστης στο document 5
  - Όπως θα δούμε λίγο παρακάτω…!

Πλεονεκτήματα συστημάτων content-based filtering

- Δεν είναι απαραίτητα δεδομένα από άλλους χρήστες
  - Μόνο τα rankings του χρήστη για τον οποίον προορίζεται το recommendation αρκούν
- Μπορούν να δώσουν recommendations σε χρήστες με προτιμήσεις που δεν είναι δημοφιλείς
- Μπορούν να προτείνουν νέα items καθώς και items που δεν είναι δημοφιλή
- Μπορούν να εξηγούν για ποιον λόγο πρότειναν κάποιο item, παρουσιάζοντας μια λίστα με τα χαρακτηριστικά του item που οδήγησαν στο recommendation

Μειονεκτήματα συστημάτων content-based filtering (1/2)

- Πρέπει να γνωρίζουν το content των items
  - Αυτό απαιτεί indexing (είτε manual είτε αυτόματο)
  - Τα items δεν μπορούν πάντα να περιγραφούν πλήρως από μια σειρά χαρακτηριστικών
- “User cold-start” problem
  - Πρέπει να μάθουν ποια χαρακτηριστικά του περιεχόμενου των items είναι σημαντικά για ένα νέο χρήστη και αυτό απαιτεί χρόνο
- Τι συμβαίνει στην περίπτωση που οι προτιμήσεις κάποιου χρήστη αλλάξουν?
  - Το σύστημα εξακολουθεί να προτείνει items με βάση τις παλιές προτιμήσεις

Μειονεκτήματα συστημάτων content-based filtering (2/2)

- Over-specialization
  - Ο χρήστης περιορίζεται σε similar items
  - Δεν δίνεται η δυνατότητα σε έναν χρήστη να ανακαλύψει items τα οποία μπορεί να του αρέσουν ακόμα και αν δεν συμφωνούν με τις προτιμήσεις που έχει δηλώσει ο ίδιος (explicit feedback) ή έχει ανακαλύψει το σύστημα για αυτόν (implicit feedback)

- Δεν εκμεταλλεύονται κριτικές για items που προέρχονται από άλλους χρήστες

Collaborative Filtering συστήματα

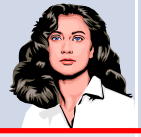
- Βασίζονται σε communities χρηστών
- Προτείνουν items σε έναν χρήστη με βάση τις προτιμήσεις άλλων χρηστών
  - Δεν απαιτείται ανάλυση του content
- Οι χρήστες δίνουν τις προτιμήσεις τους είτε άμεσα είτε έμμεσα (explicit ή implicit feedback)
- Τρεις βασικές κατηγορίες
  - User-based collaborative filtering
  - Item-based collaborative filtering
  - Υβριδικά

User-based Collaborative Filtering

- Βασική ιδέα:
  - Οι χρήστες που συμφώνησαν στο παρελθόν (σχετικά με items) τείνουν να συμφωνήσουν πάλι στο μέλλον.
- Για να προβλέψουν το opinion ενός χρήστη (ενεργός χρήστης) για ένα item (για το οποίο δεν έχει φυσικά ακόμα εκφράσει άποψη) χρησιμοποιούν τα opinions των similar users
- Το similarity ανάμεσα σε χρήστες βασίζεται στο κατά πόσο τα opinions που είχαν στο παρελθόν για άλλα items ήταν παρόμοια
  - γειτονιά του ενεργού χρήστη: χρήστες με παρόμοιες βαθμολογίες και προτιμήσεις για έναν μεγάλο αριθμό items που έχουν καταναλωθεί από τον ενεργό χρήστη

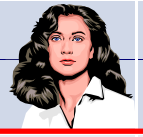
Παράδειγμα

- Ratings των users για διάφορα items
  - Ποιο είναι το προβλεπόμενο rating του user 1 για το item 3?

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 	8	1	?	2	7
User 2 	2	-	5	7	5
User 3 	5	4	7	4	7
User 4 	7	1	7	3	8
User 5 	1	7	4	6	5
User 6 	8	3	8	3	7

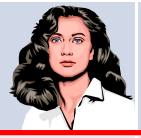
Similarity ανάμεσα σε χρήστες

- Πόσο similar είναι οι users 1 και 2?
- Πόσο similar είναι οι users 1 και 4?
- Πως θα υπολογίζατε το similarity?

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 	8	1	?	2	7
User 2 	2	-	5	7	5
User 4 	7	1	7	3	8

Υπολογισμός του similarity με Ευκλείδεια Απόσταση

Similarity των user1 και user2

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 	8	1	?	2	7
User 2 	2	-	5	7	5

- Λάβετε υπ' όψιν σας μόνο τα items που έχουν κάνει rate και οι δύο χρήστες
- Για κάθε item:
  - Υπολογίστε την διαφορά ανάμεσα στα ratings των δυο χρηστών
  - Υψώνουμε στο τετράγωνο και προσθέτουμε με τα προηγούμενα
- Στο τέλος διαιρούμε με τη ρίζα του αθροίσματος των τετραγώνων

For Item j rated by User 1 and User 2:

sum = sum+ (rating (User 1, Item j) – rating (User 2, Item j)) ^{2}

Similarity (0,1) = $\frac{1}{1+\sqrt{sum}}$

Υπολογισμός του similarity με Συσχέτιση Pearson

- Πως λαμβάνουμε υπόψη το ότι ο ένας είναι γενικά πιο αυστηρός;
- Υπολογίζουμε το άθροισμα των ratings και του τετραγώνου των ratings για κάθε χρήστη
  - Sum1=sum(ratings user1), Sum2=sum(ratings user2)
  - Sum1Sq=sum[(ratings user1)^2], Sum2Sq=sum[(ratings user2)^2]
- Υπολογίζουμε το άθροισμα των γινομένων των ratings για τα n αντικείμενα
  - pSum=sum((rating user1 for item i)*(rating user2 for item i))

$$num = pSum - \frac{sum1 \cdot sum2}{n}$$

$$similarity = \frac{num}{den}$$

$$den = \sqrt{\left(\frac{sum1Sq - sum1^2}{n}\right) \cdot \left(\frac{sum2Sq - sum2^2}{n}\right)}$$

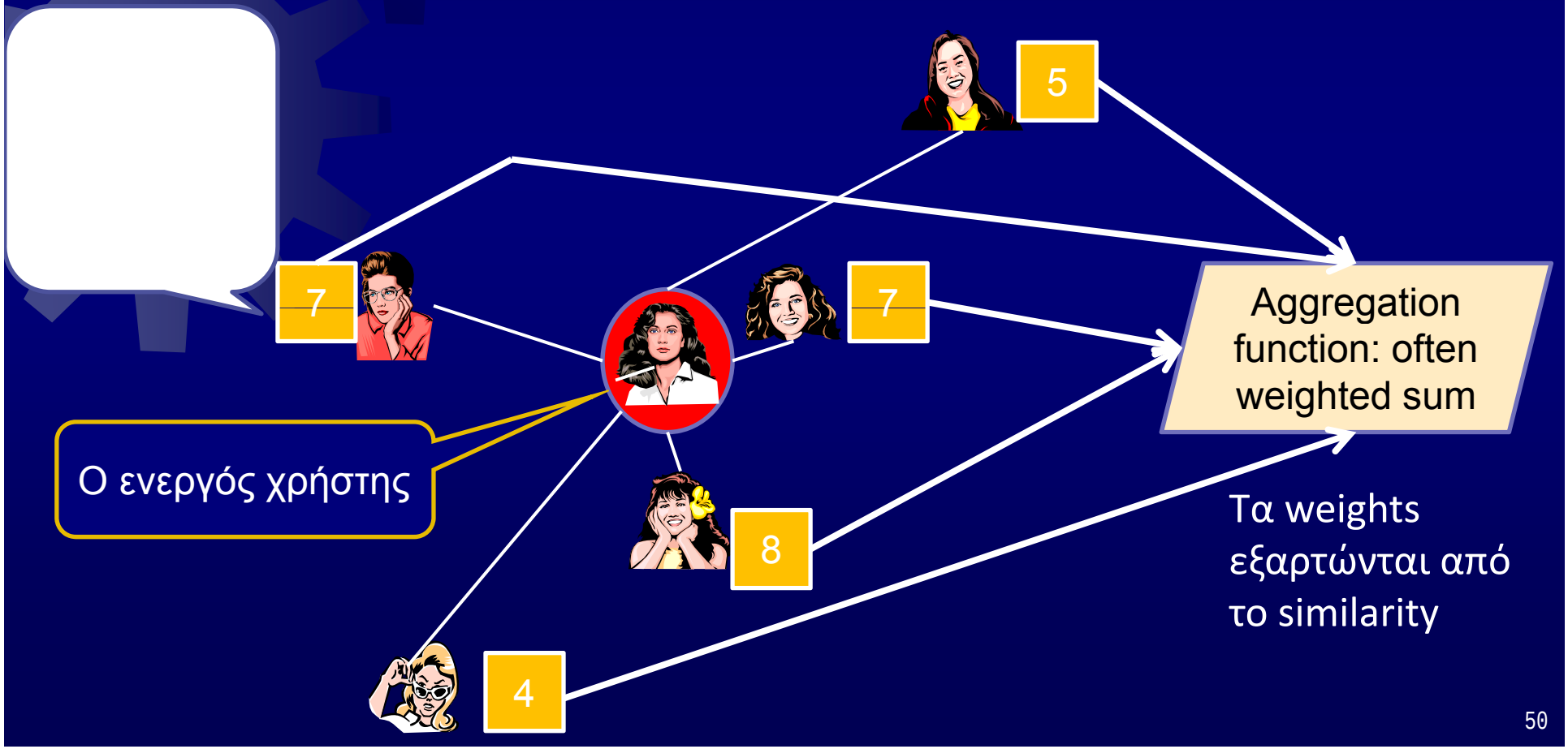
Υπολογισμός του similarity με cosine similarity coefficient

$$\mathrm{sim}(u,u') = \frac{\left(\sum_{i \in I(u,u')} R(u,i) R(u',i)\right)}{\left(\sqrt{\sum_{i \in I(u,u')} R(u,i)^2} \sqrt{\sum_{i \in I(u,u')} R(u',i)^2}\right)}$$

- Όπου R(u, i) είναι το rating που έδωσε ο user u για το item i.
- και I(u, u') είναι το σύνολο των items που έχουν αξιολογηθεί από τον user u και u'

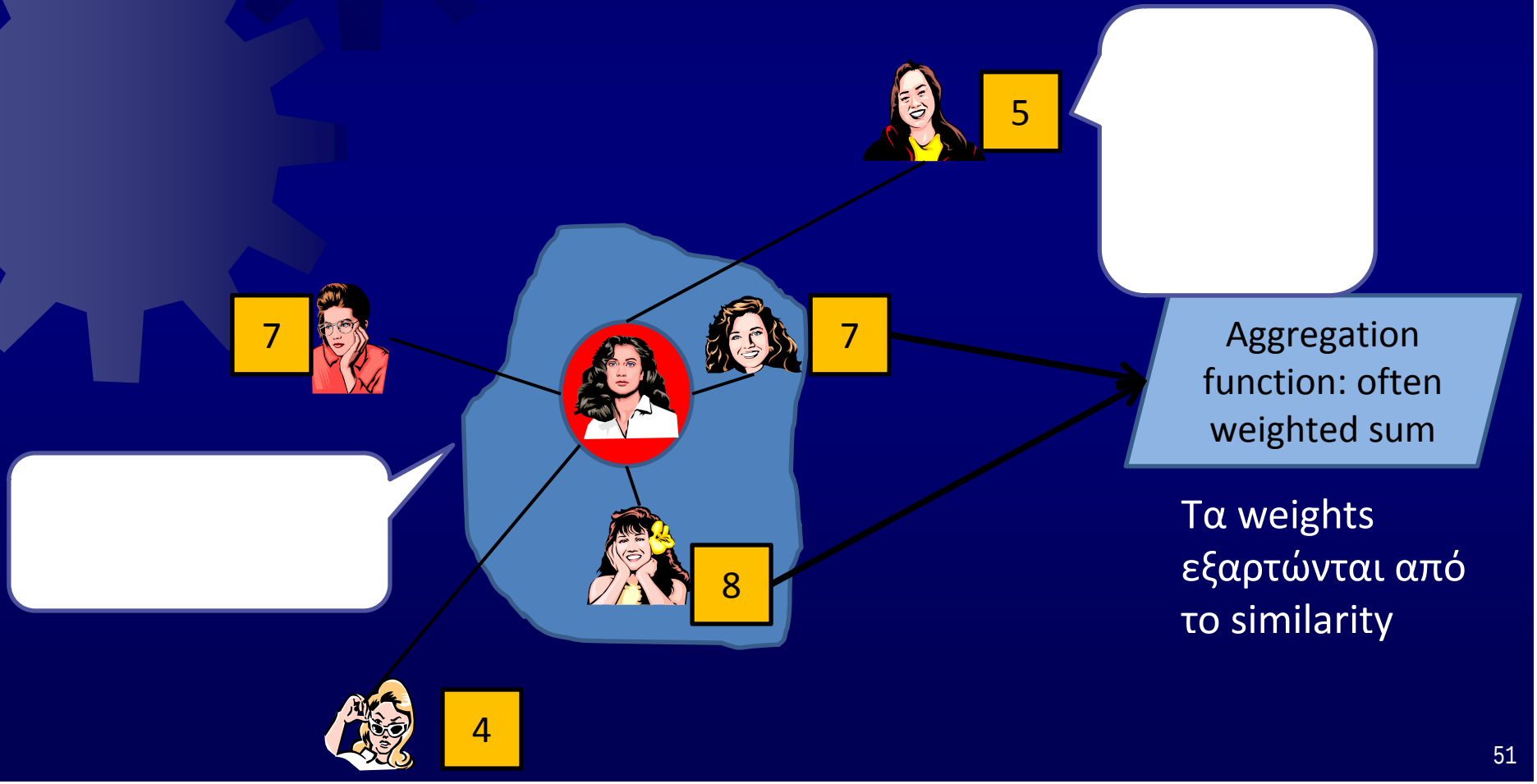
Υπολογισμός predicted rating του ενεργού χρήστη

- Η απόσταση ανάμεσα στον ενεργό χρήστη και τους άλλους 5 χρήστες είναι ανάλογη με το πόσο similar είναι
- **1η μέθοδος**: Χρησιμοποιούνται τα ratings **ΟΛΩΝ** των άλλων χρηστών για να υπολογίζουμε το predicted rating, μέσω μιας aggregation function



2η μέθοδος: K-Nearest-Neighbour

- **2η μέθοδος**: Χρησιμοποιούνται τα ratings **<u>των πιο κοντινών γειτόνων</u>** του ενεργού χρήστη (δυο σε αυτό το παράδειγμα) για να υπολογίζουμε το predicted rating για αυτόν, μέσω μιας aggregation function



Aggregation function

- Weighted sum:

$$R(u,i) = z \sum_{u' \in N(u)} \text{sim}(u,u') \cdot R(u',i)$$

- Adjusted weighted sum

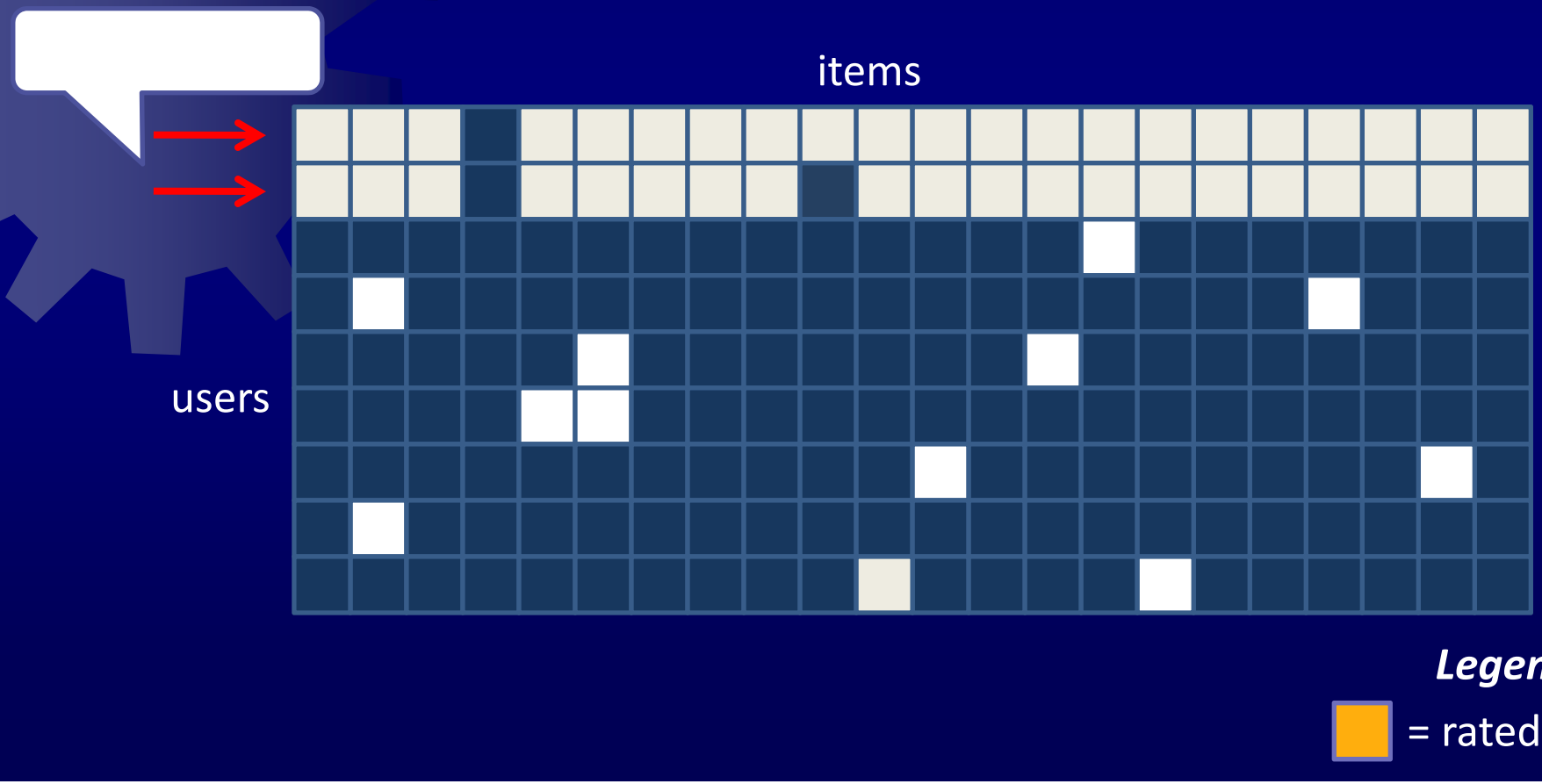
$$R(u,i) = \overline{R(u)} + z \sum_{u' \in N(u)} \text{sim}(u,u') \cdot \left(R(u',i) - \overline{R(u')} \right)$$

- όπου $\overline{R(u)}$ είναι το μέσο rating του user u και N(u) το σετ των πιο κοντινών γειτόνων του user u.
- Z είναι παράμετρος που συνήθως δίνεται από:

$$z = \frac{1}{\sum_{u' \in N(u)} \left| \text{sim}(u,u') \right|}$$

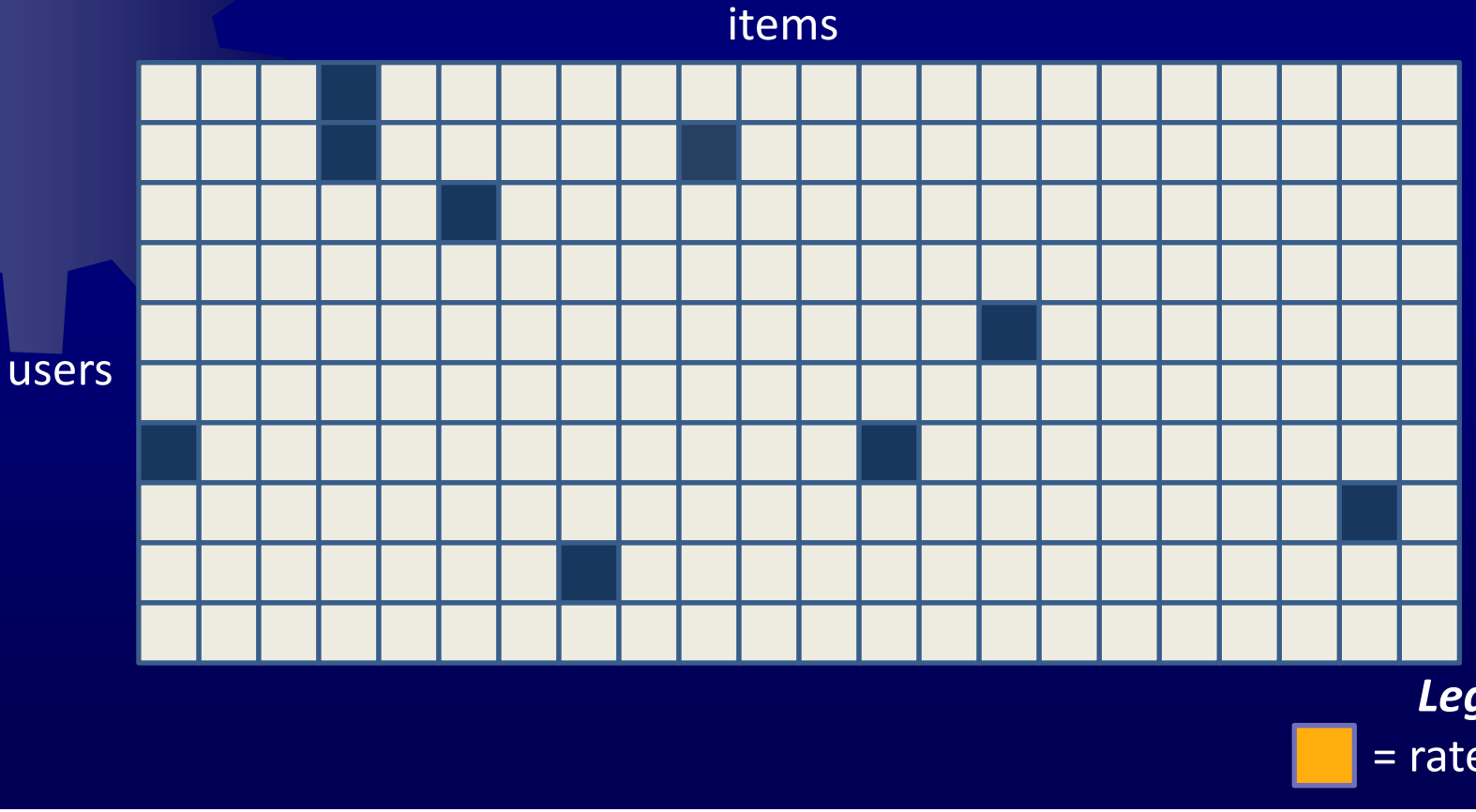
User-based Collaborative Filtering: Προβλήματα (1/3)

- User Cold-Start problem
  - Οι νέοι χρήστες (ίσως και κάποιοι παλιοί) δεν έχουν κάνει πολλά ratings για να αποφασίσουμε με ποιους άλλους χρήστες είναι similar



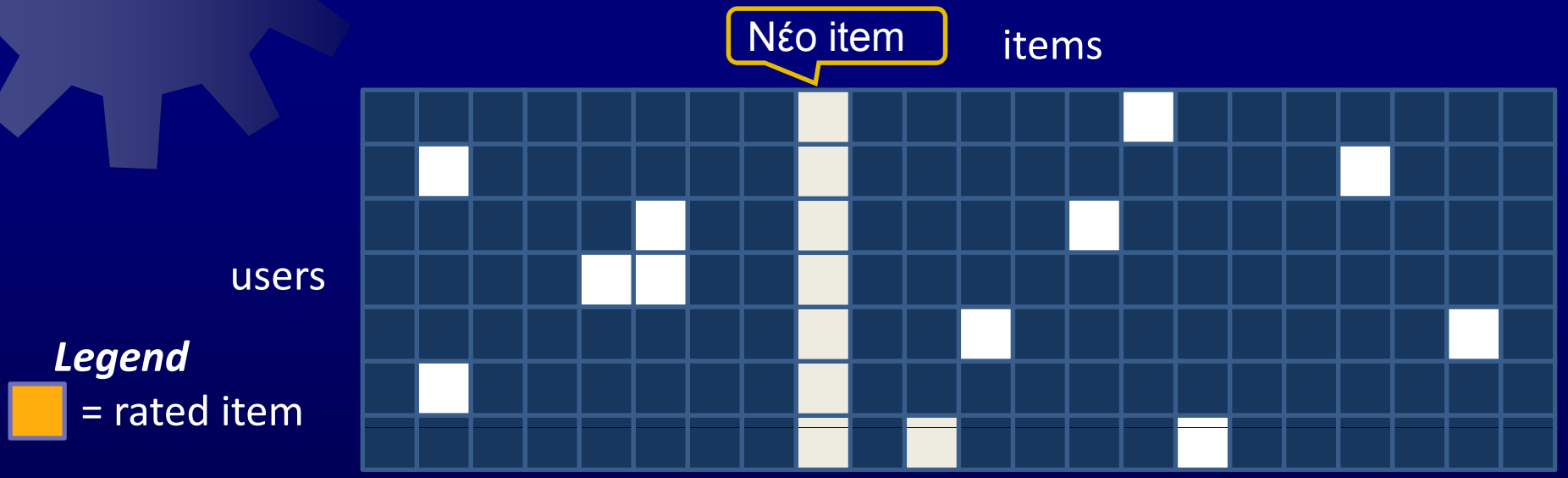
User-based Collaborative Filtering: Προβλήματα (2/3)

- Σποραδικότητα (sparsity) των ratings
  - Αν το σύνολο των items είναι μεγάλο, οι χρήστες μπορεί να έχουν κάνει rate μόνο λίγα items (είναι δύσκολο να βρεθούν similar users)



User-based Collaborative Filtering: Προβλήματα (3/3)

- Item Cold-Start problem
  - Το σύστημα δεν μπορεί να προβλέψει ratings για ένα νέο item, μέχρι να το αξιολογήσουν κάποιοι users
    - Αυτό το πρόβλημα ΔΕΝ υπήρχε στο content-based filtering
- Scalability
  - Με εκατομμύρια χρήστες και items οι υπολογισμοί μπορεί να γίνουν πολύ αργοί



Demographic Recommenders

- Για να προβλέψουν το opinion ενός χρήστη για ένα item χρησιμοποιούν τα opinions των similar users
  - Ίδια ιδέα με τα συστήματα user-based Collaborative Filtering

- Σε αυτά τα συστήματα όμως, το similarity ανάμεσα στους χρήστες βασίζεται στα δημογραφικά στοιχεία τους (stereotypes)
  - Π.χ. ηλικία, εκπαίδευση, φύλο κτλ
  - Οι χρήστες πρέπει να δίνουν τα δημογραφικά τους στοιχεία

Item-based Collaborative Filtering

- **Βασική ιδέα**: Ένας χρήστης είναι πιθανό να έχει την ίδια γνώμη για similar items
  - Ίδια ιδέα με το Content-Based Filtering
- Το **similarity** ανάμεσα στα items βασίζεται στον τρόπο που τα έχουν κάνει rate άλλοι χρήστες
  - Διαφέρει από το content-based, που χρησιμοποιεί τα χαρακτηριστικά (content) των items
- **Πλεονεκτήματα** σε σχέση με το user-based CF:
  - Αποτρέπει το User Cold-Start problem
  - Βελτιώνει το scalability (το similarity ανάμεσα σε items είναι πιο stable σε σχέση με το similarity ανάμεσα σε users)

Παράδειγμα

- Ποιο είναι το προβλεπόμενο rating του user 1 για το item 3?

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1 	8	1	?	2	7
User 2 	2	-	5	7	5
User 3 	5	4	7	4	7
User 4 	7	1	7	3	8
User 5 	1	7	4	6	5
User 6 	8	3	8	3	7

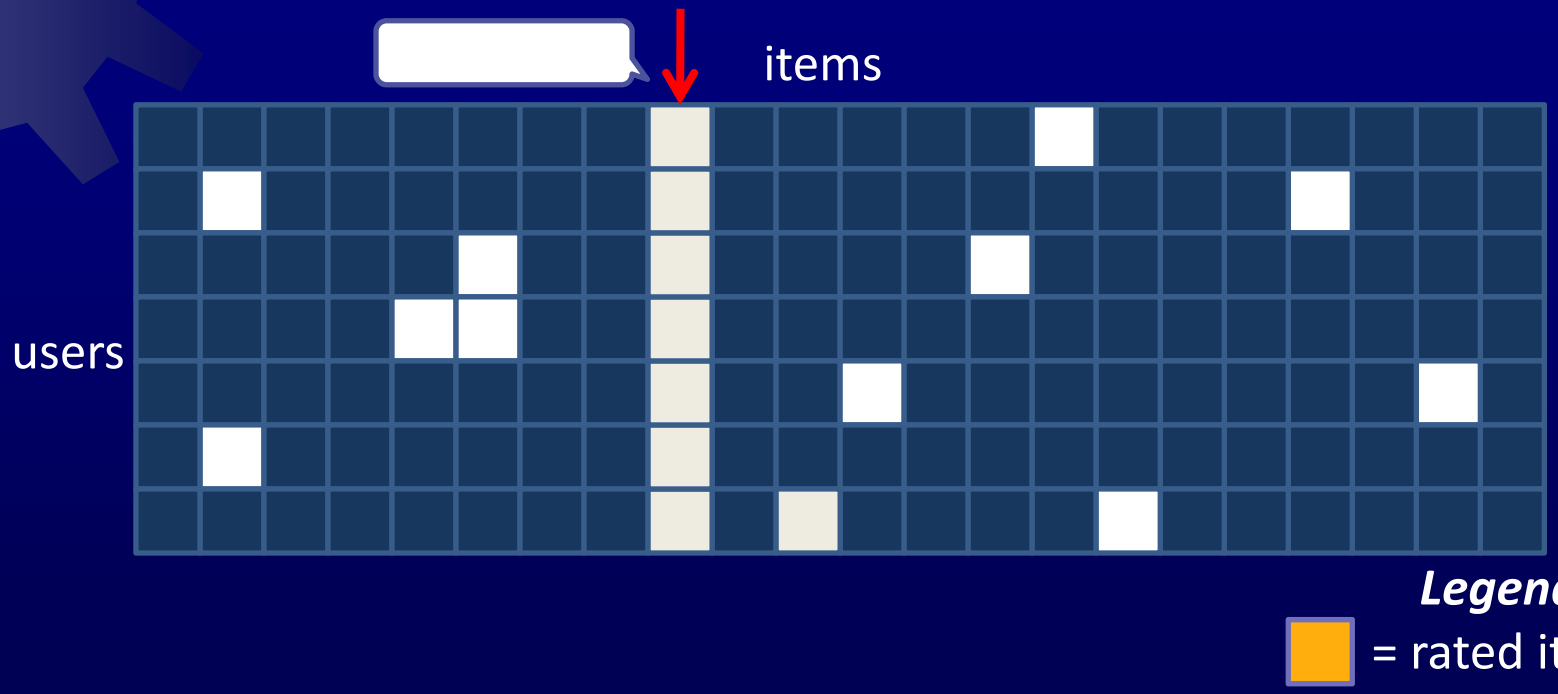
Similarity ανάμεσα σε items

Item 3	Item 4	Item 5
?	2	7
5	7	5
7	4	7
7	3	8
4	6	5
8	3	7

- Πόσο similar είναι τα items 3 και 4?
- Πόσο similar είναι τα items 3 και 5?
- Πως θα υπολογίζατε το similarity?

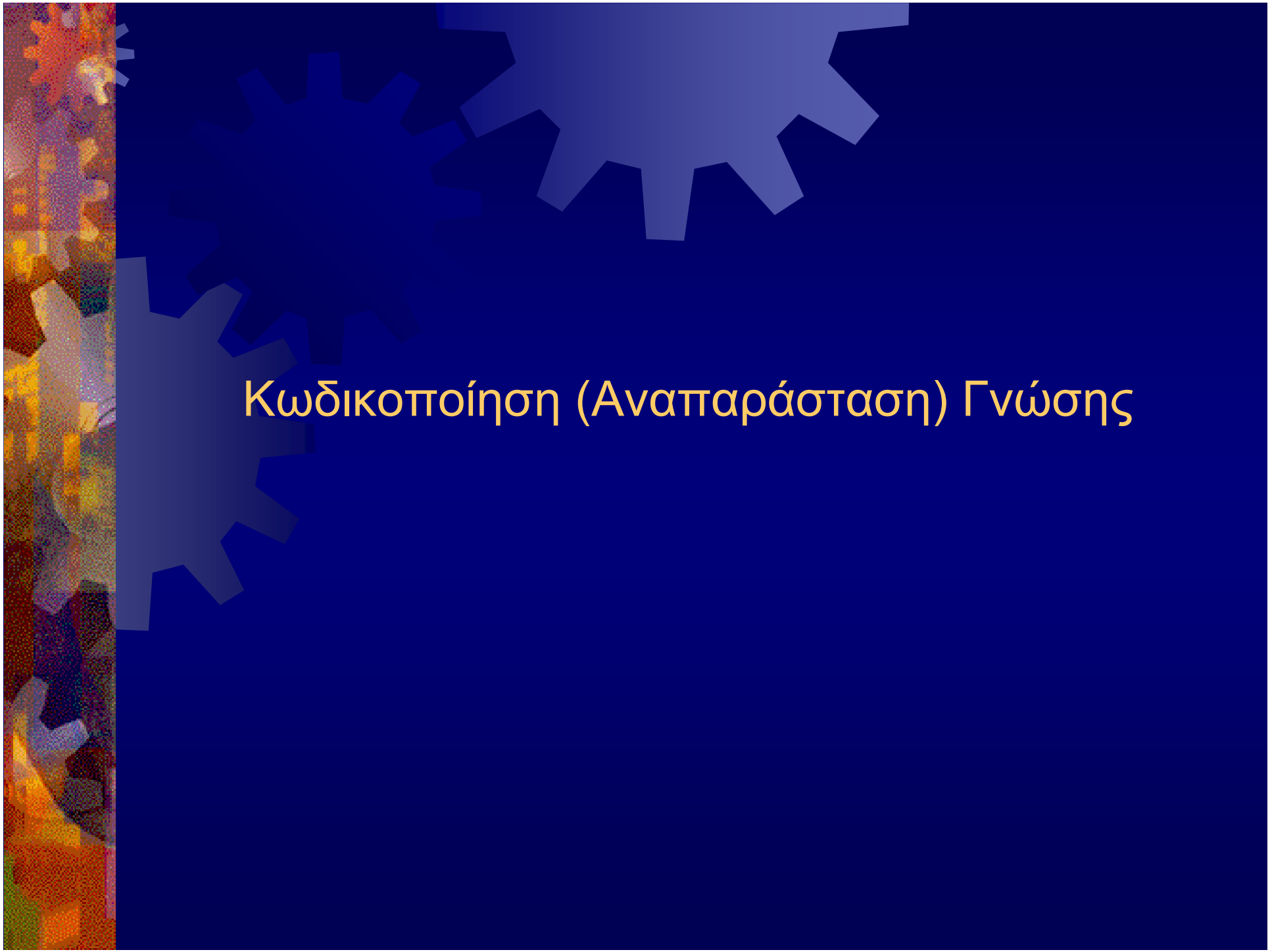
Item-based Collaborative Filtering: Προβλήματα

- Item Cold-Start problem
  - Το σύστημα δεν μπορεί να βρει ποια items είναι similar, μέχρι να το αξιολογήσουν κάποιοι users
  - Content-based: Δεν υπήρχε αυτό το πρόβλημα
  - User-based CF: υπήρχε και εκεί αλλά εδώ είναι μεγαλύτερο



Υβριδικά Συστήματα

- Συνδυασμός Content-based και Collaborative Filtering
- Ή συνδυασμός User-based και Item-based Collaborative Filtering
- Για να αντιμετωπιστούν τα προβλήματα που έχει η κάθε τεχνική
- Some ways to make a Hybrid
  - Weighted. Ratings of several recommendation techniques are combined together to produce a single recommendation
  - Switching. The system switches between recommendation techniques depending on the current situation
  - Mixed. Recommendations from several different recommenders are presented simultaneously (e.g. Amazon)
  - Cascade. One recommender refines the recommendations given by another


Κωδικοποίηση (Αναπαράσταση) Γνώσης

Κωδικοποίηση γνώσης στον κύκλο ζωής συστήματος ΔΓ

Τι είναι κωδικοποίηση γνώσης;

- Οργάνωση και αναπαράσταση γνώσης πριν από την πρόσβαση σε αυτήν και χρήση αυτής
- Μετατροπή μέρους της άρρητης γνώσης σε ρητή (σε επαναχρησιμοποιήσιμη μορφή)
- Μετατροπή μη-τεκμηριωμένης γνώσης σε τεκμηριωμένη
- Μετατροπή εταιρικής γνώσης σε μορφή προσβάσιμη, ορατή και χρησιμοποιήσιμη για λήψη αποφάσεων

Αναπαράσταση με Κανόνες

- Πολύ πρακτικός τρόπος αναπαράστασης για την εξαγωγή συμπερασμάτων
- Αποτελούν τη βάση πολλών συστημάτων γνώσης (knowledge systems)
- Γενικά Πλεονεκτήματα:
  - Κοντά στην ανθρώπινη γνώση
  - Επάρκεια συνεπαγωγών
- Συγκεκριμένα Πλεονεκτήματα:
  - Modularity: Κάθε κανόνας ορίζει ένα μικρό, (σχεδόν) ανεξάρτητο τμήμα της γνώσης
  - Incrementability: Μπορούν να προστεθούν νέοι κανόνες (σχεδόν) ανεξάρτητα από τους υπάρχοντες
  - Modifiability: Οι υπάρχοντες κανόνες μπορούν να αλλάξουν (σχεδόν) ανεξάρτητα από τους υπόλοιπους

Παράδειγμα Αναπαράστασης με Κανόνες

- Συνεπαγωγικός Κανόνας

  IF ο εκτυπωτής τυπώνει σωστά AND
  τα χρώματα δε τυπώνονται σωστά
  THEN έχει τελειώσει το έγχρωμο μελάνι

- Κανόνας Παραγωγής

  IF ο εκτυπωτής τυπώνει σωστά AND
  τα χρώματα δε τυπώνονται σωστά
  THEN αλλάξτε την κεφαλή με το έγχρωμο μελάνι

- Ενεργός Κανόνας

  ON εκτύπωση
  IF τα χρώματα δε τυπώνονται σωστά
  THEN αλλάξτε την κεφαλή με το έγχρωμο μελάνι

Πίνακες Λήψης Αποφάσεων

- Χρησιμοποιείται ως φύλλο εργασίας, διαμερισμένο σε λίστα συνθηκών (και τιμές αυτών) και λίστα συμπερασμάτων
- Συνθήκες αντιστοιχίζονται με συμπεράσματα

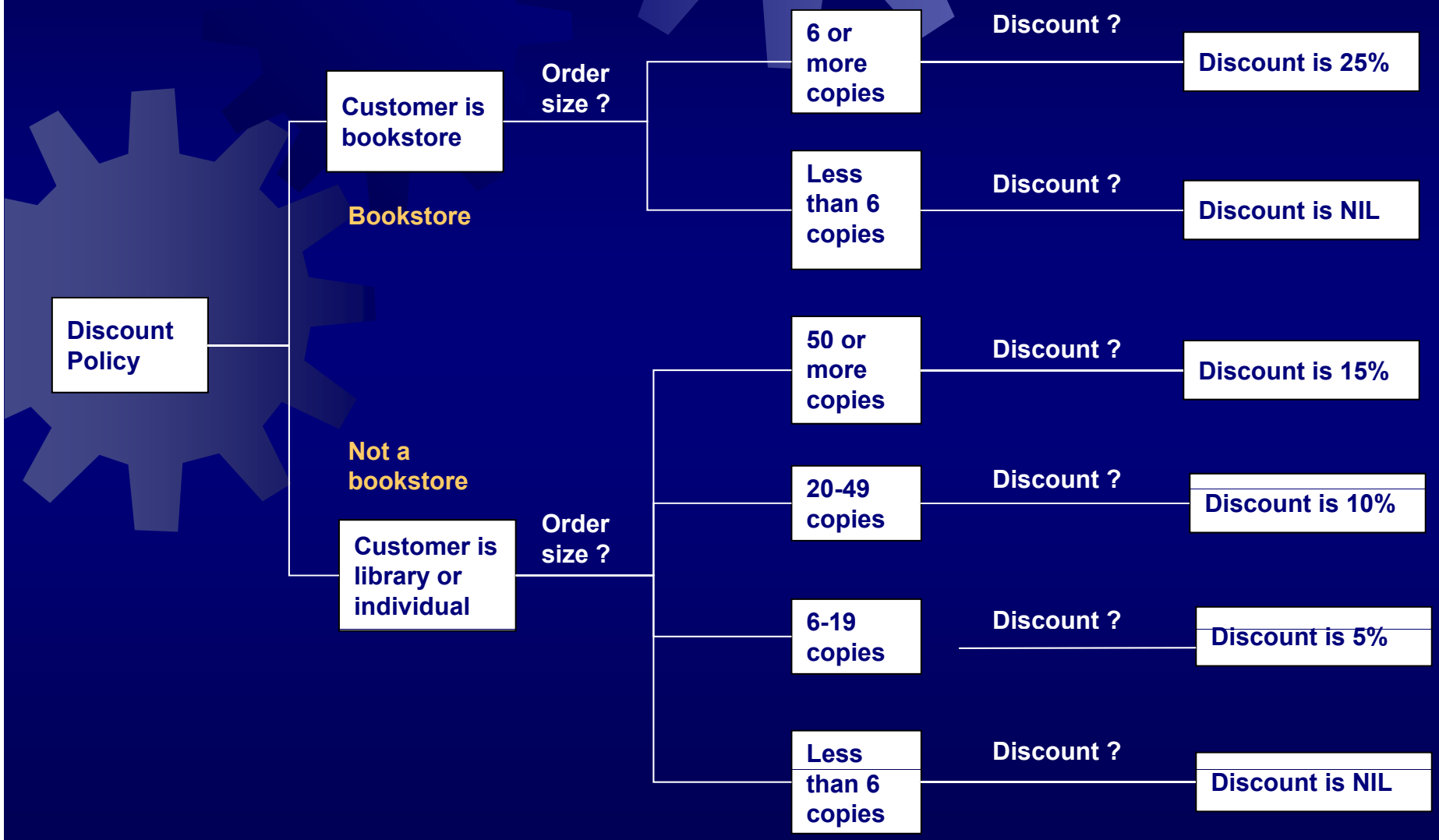
Ενδεικτικός Πίνακας Λήψης Αποφάσεων

	Λίστα συνθηκών	Τιμές					
		1	2	3	4	5	6
IF (συνθήκη)	Customer is bookstore	Y	Y	N	N	N	N
	Order size > 6 copies	Y	N	N	N	N	N
	Customer is librarian/individual			Y	Y	Y	Y
	Order size 50 copies or more			Y	N	N	N
	Order size 20-49 copies				Y	N	N
	Order size 6-19 copies					Y	N
THEN (ενέργεια)	Allow 25% discount	X					
	Allow 15% discount			X			
	Allow 10% discount				X		
	Allow 5% discount					X	
	Allow no discount		X				X
	Λίστα ενεργειών	Τιμές					

Δένδρα Λήψης Αποφάσεων

- Ιεραρχικά δομημένο σημασιολογικό δίκτυο (semantic network)
- Αποτελείται από κόμβους (αναπαριστούν στόχους ή αποτελέσματα αποφάσεων) και συνδέσεις (αναπαριστούν συνθήκες)
- Διατρέχεται από αριστερά προς τα δεξιά (απαρχή αριστερά)
- Δένδρα Λήψης Αποφάσεων παρέχουν δυνατότητα απεικόνισης λογικής αλληλουχίας ενεργειών για επίτευξη στόχου

Παράδειγμα Δένδρου Λήψης Αποφάσεων



Πλαίσια απεικόνισης

- Απεικονίζουν γνώση που αφορά μία περιοχή
- Εξετάζουν συνδυασμό επεξηγηματικής (declarative) και επιχειρησιακής (operational) γνώση που διευκολύνει κατανόηση περιοχής υπό εξέταση
- Παρέχουν:
  - υποδοχή (slot), συγκεκριμένο αντικείμενο ή χαρακτηριστικό οντότητας, και
  - Όψη (facet),τη τιμή που αντιστοιχεί στην υποδοχή
- Όταν όλες οι υποδοχές έχουν συμπληρωθεί μα τιμές, το πλαίσιο θεωρείται συμπληρωμένο (instantiated)

Παράδειγμα πλαισίου απεικόνισης

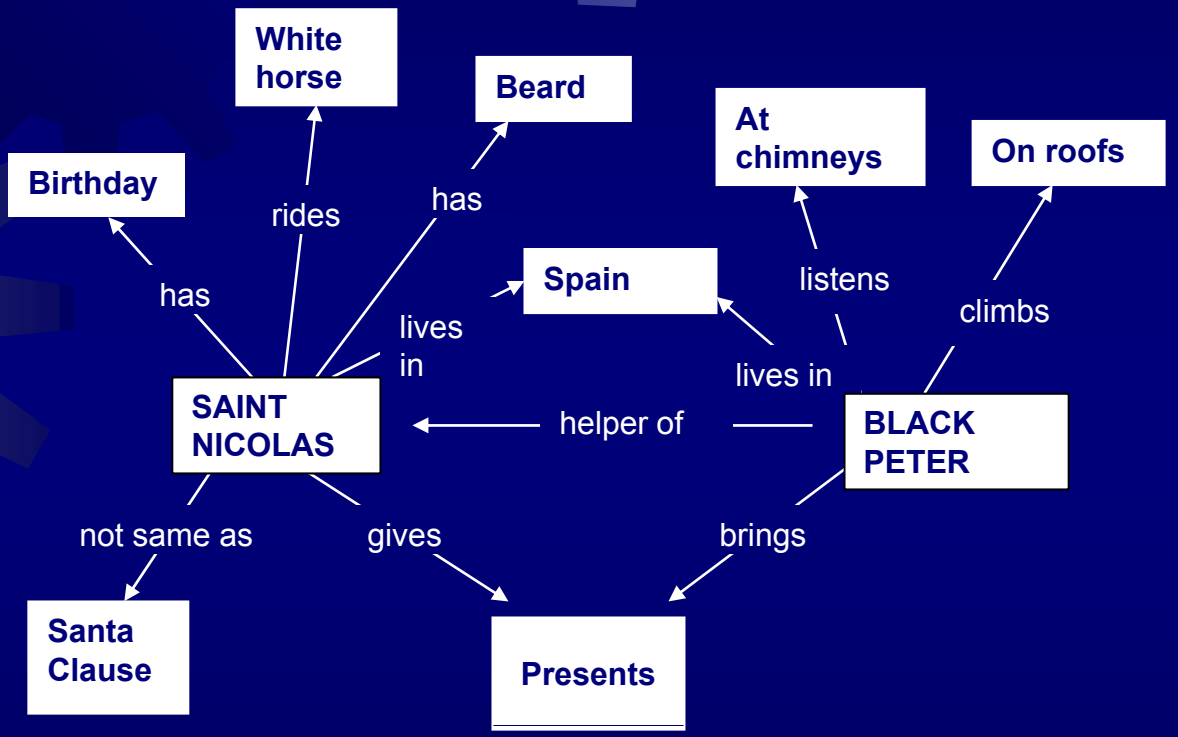
- Object: License

Slot	Facet
Instructor	Olesek
Verification	Training license
Unique feature of verification	State seal
Teaching certification	In air rescue

Χαρτογράφηση εννοιών (Concept Mapping) και Χάρτες γνώσης (Knowledge Maps)

- Δίκτυο εννοιών που αποτελείται από κόμβους και συνδέσμους
- Κόμβος απεικονίζει έννοια και σύνδεσμος τη σχέση μεταξύ εννοιών
- Αποτελεσματικός τρόπος ομαδοποίησης εννοιών χωρίς να χάνεται η μοναδικότητα.

Παράδειγμα Concept Map

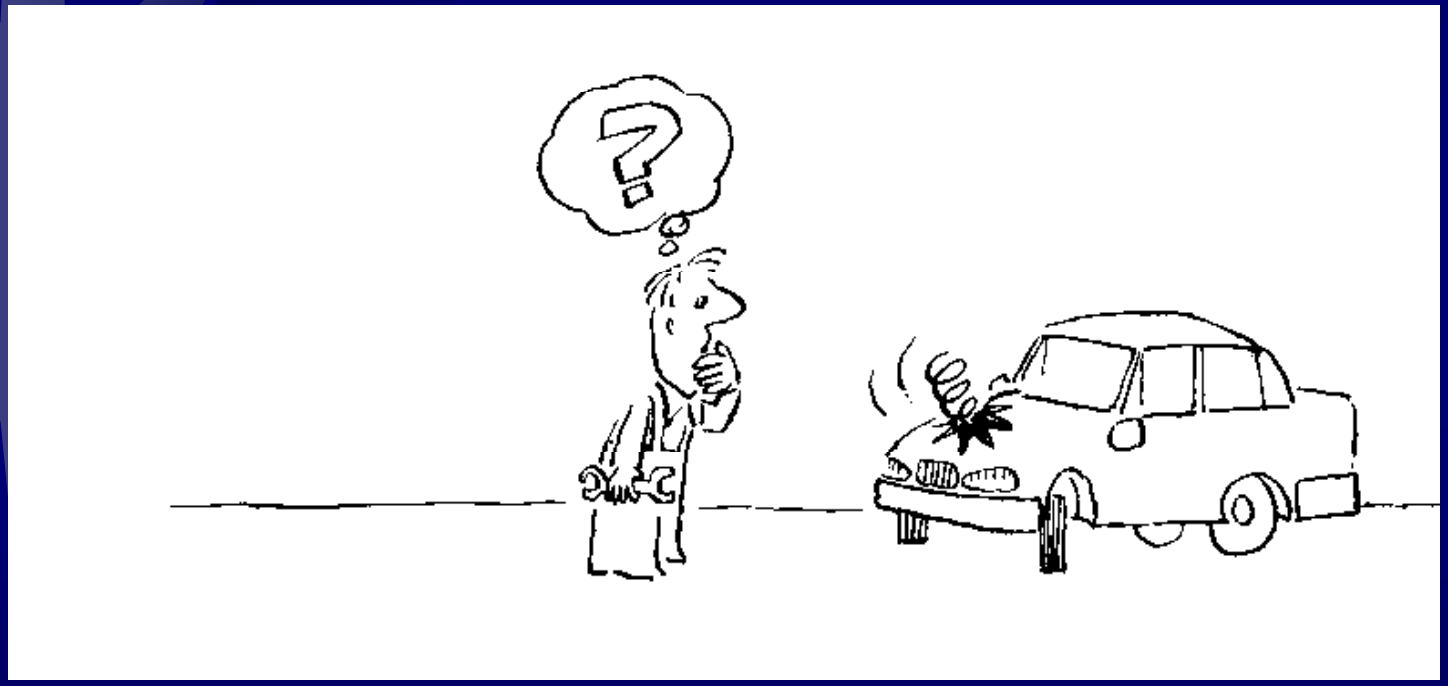


Case-Based Reasoning (CBR)

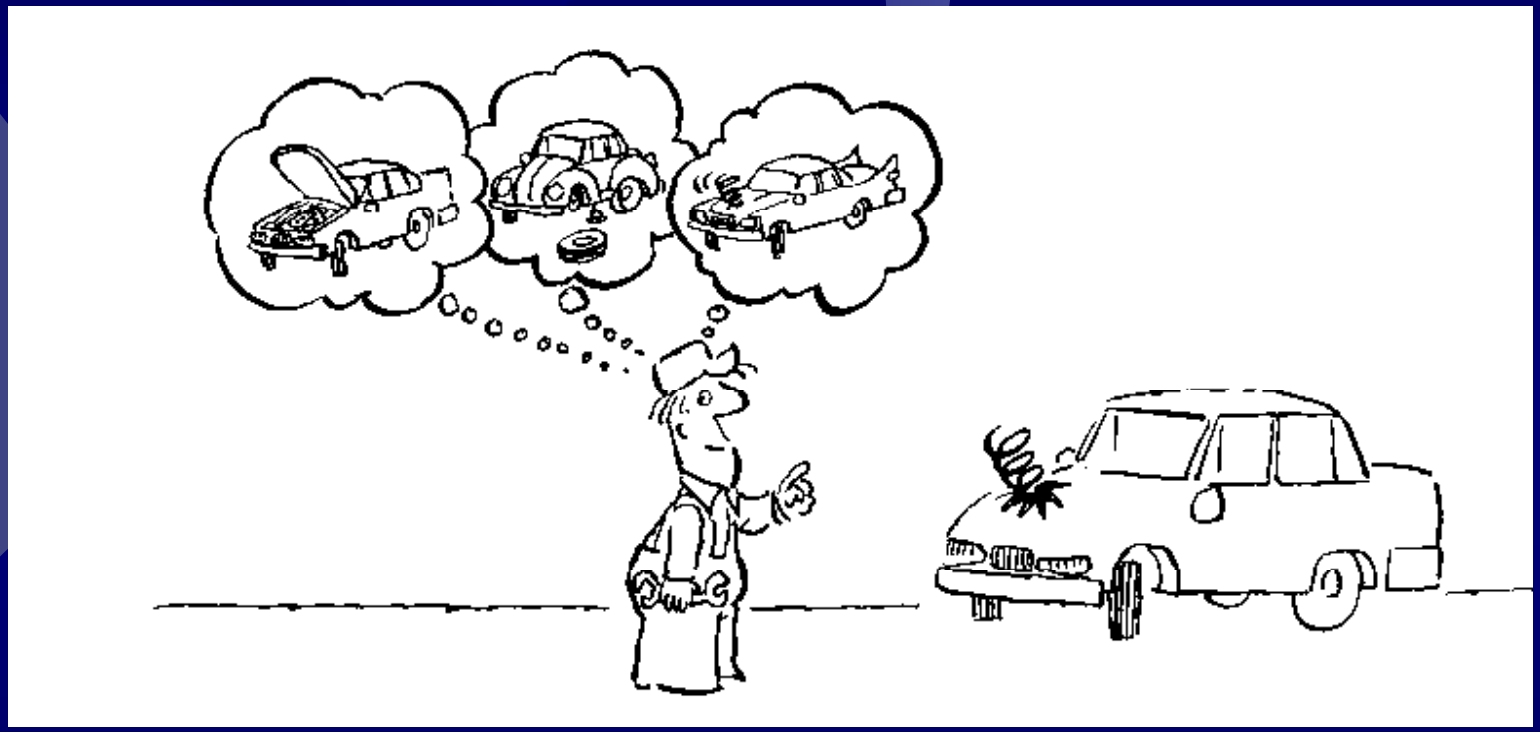
To solve a problem, remember a similar problem you have solved in the past and adapt the old solution to solve the new problem

Alex Goodall

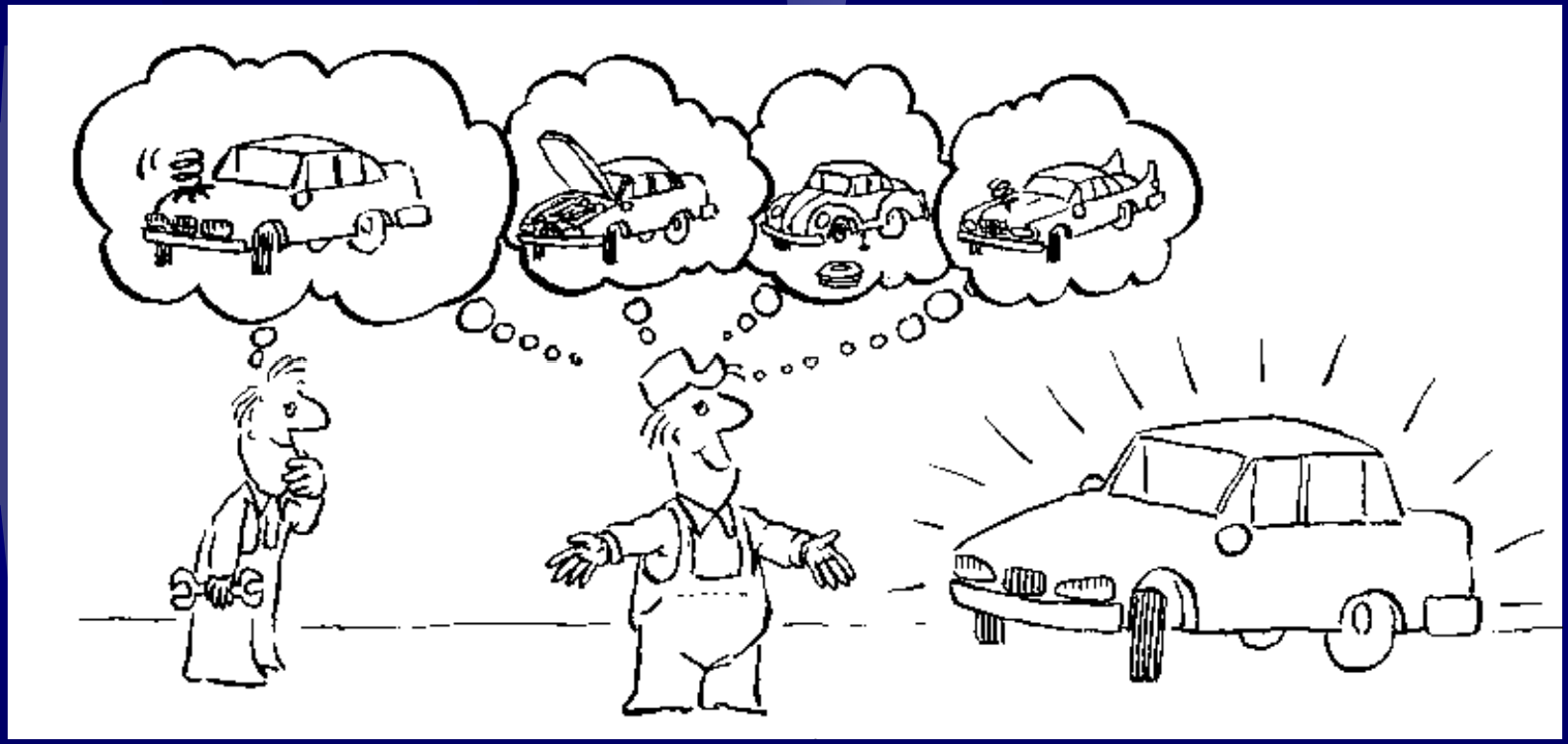
Case-Based Reasoning (CBR)



Case-Based Reasoning (CBR)



Case-Based Reasoning (CBR)

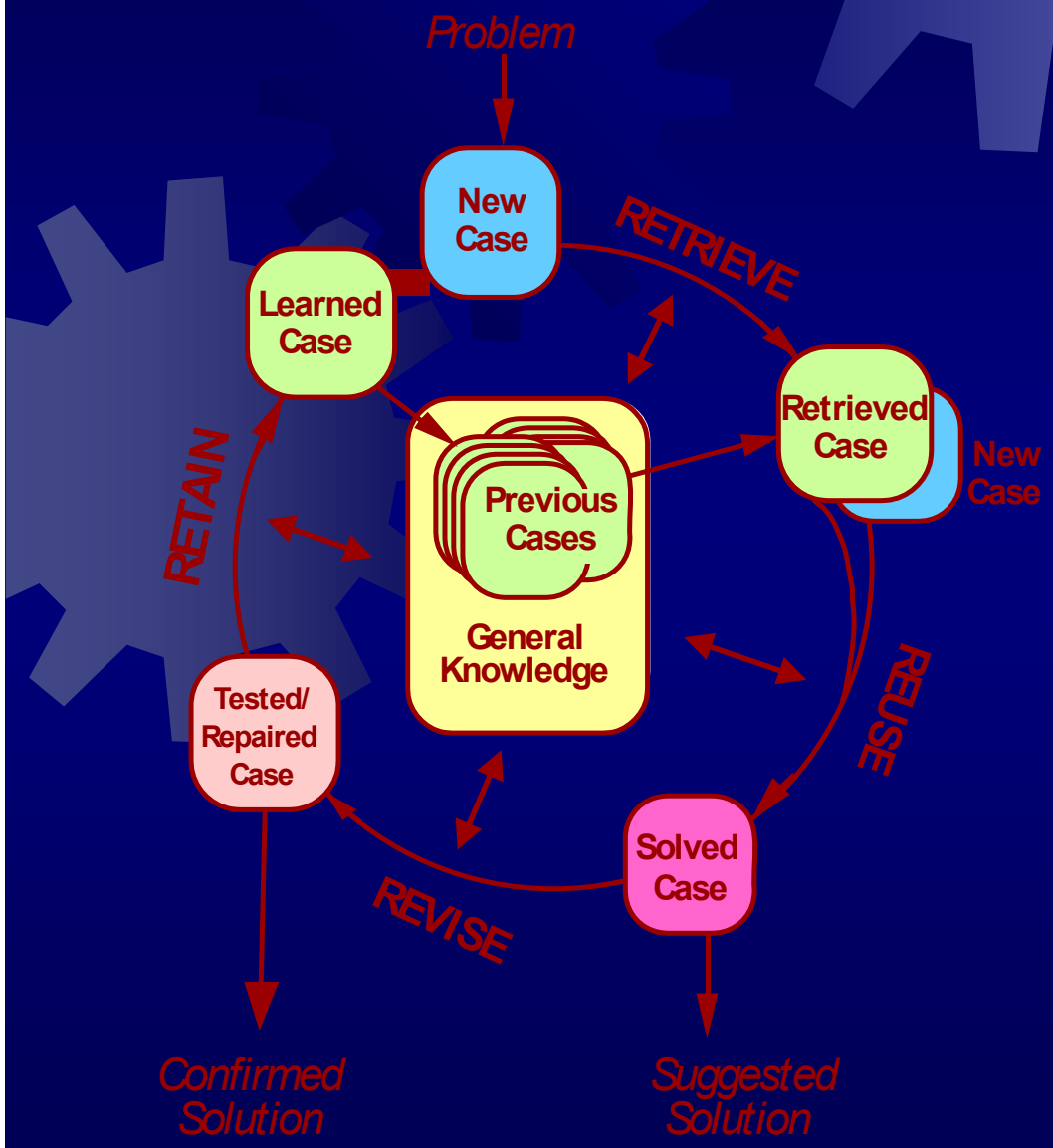


Case-Based Reasoning (CBR)

- Η **μέθοδος CBR** αποτελεί μέθοδο συλλογιστικής που βασίζεται σε σχετικές παλαιές περιπτώσεις (cases)
- Προσομοιάζει ανθρώπινη συλλογιστική βάση εμπειριών
- Στόχο αποτελεί η ανάδειξη των πιο σχετικών ιστορικών περιπτώσεων που ταιριάζουν με τη παρούσα περίπτωση
- Είναι πιο γρήγορη μέθοδος από μεθόδους rule-based
- Απαιτεί ενδελεχή αρχικό σχεδιασμό όλων των παραμέτρων σχετικότητας

Παράδειγμα CBR

Problem solving



1. Search Similar Case
2. Re-Use Solution
3. Adapt Solution
4. Store Experience
5. New Problem?

Example: FAQ

FAQ document

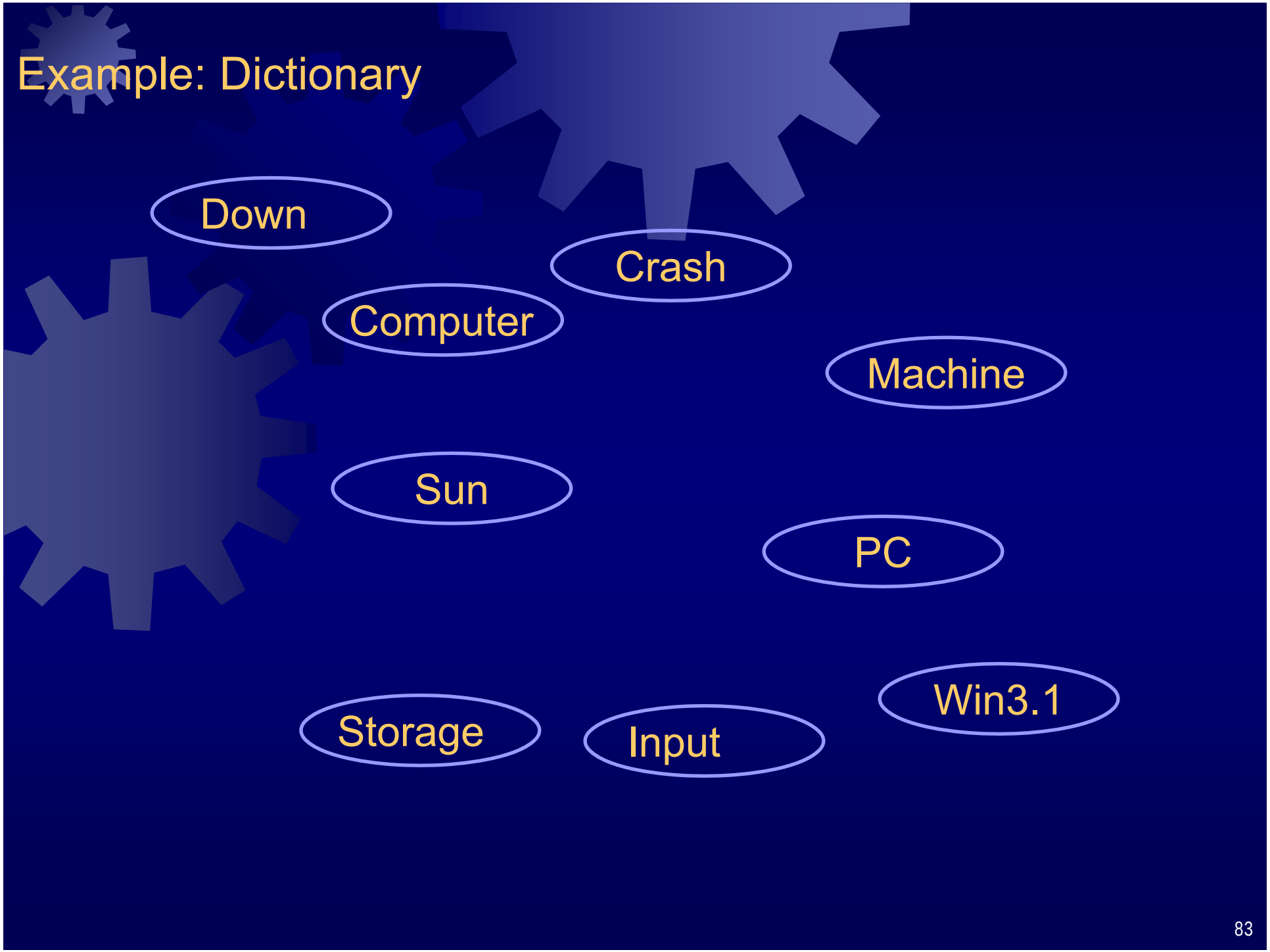
Hardware:
PC & HP DeskJet 870

Software:
Windows 95

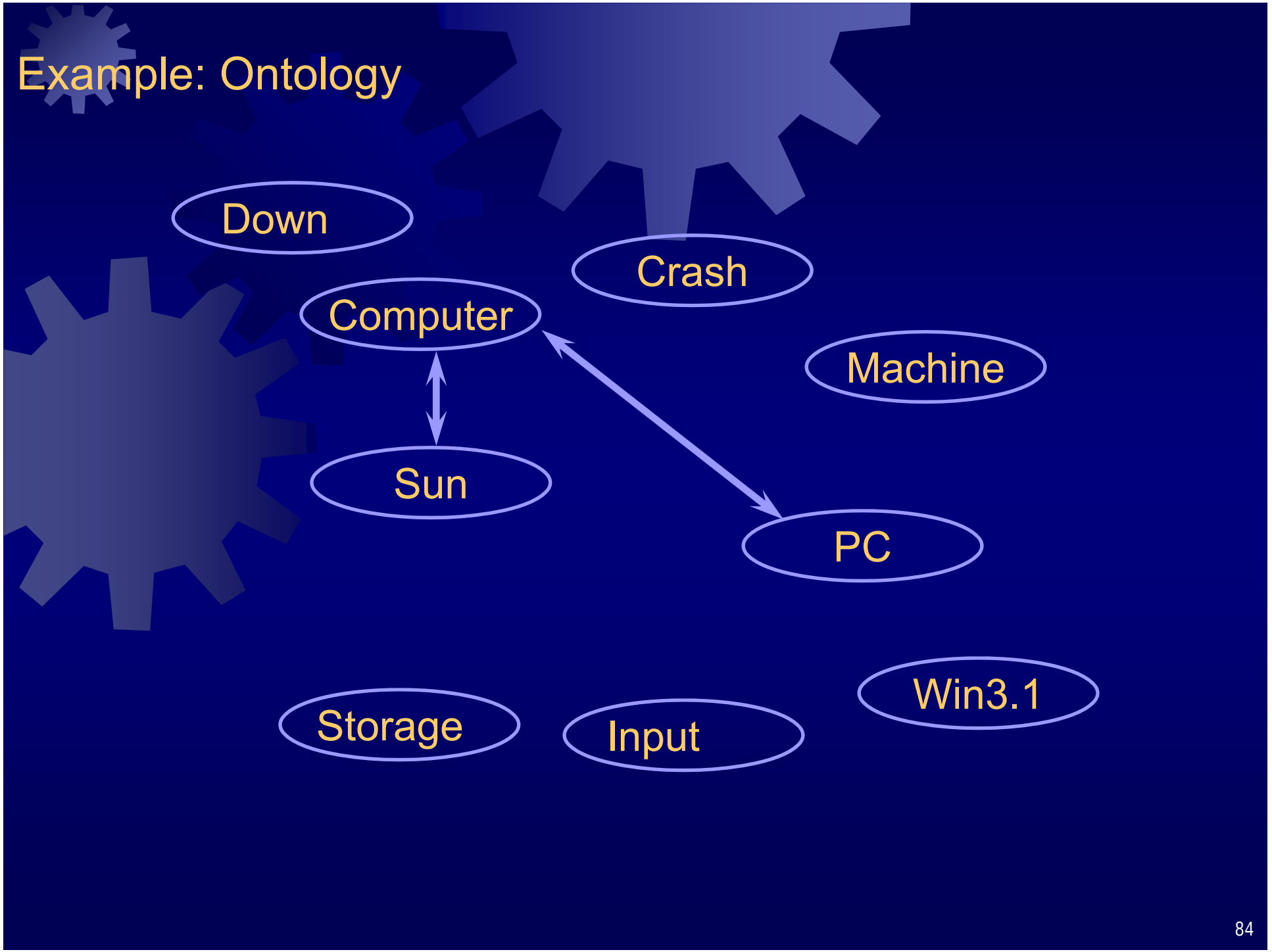
Question:
My new printer crops graphic print outs.

Answer:
load and install new printer driver

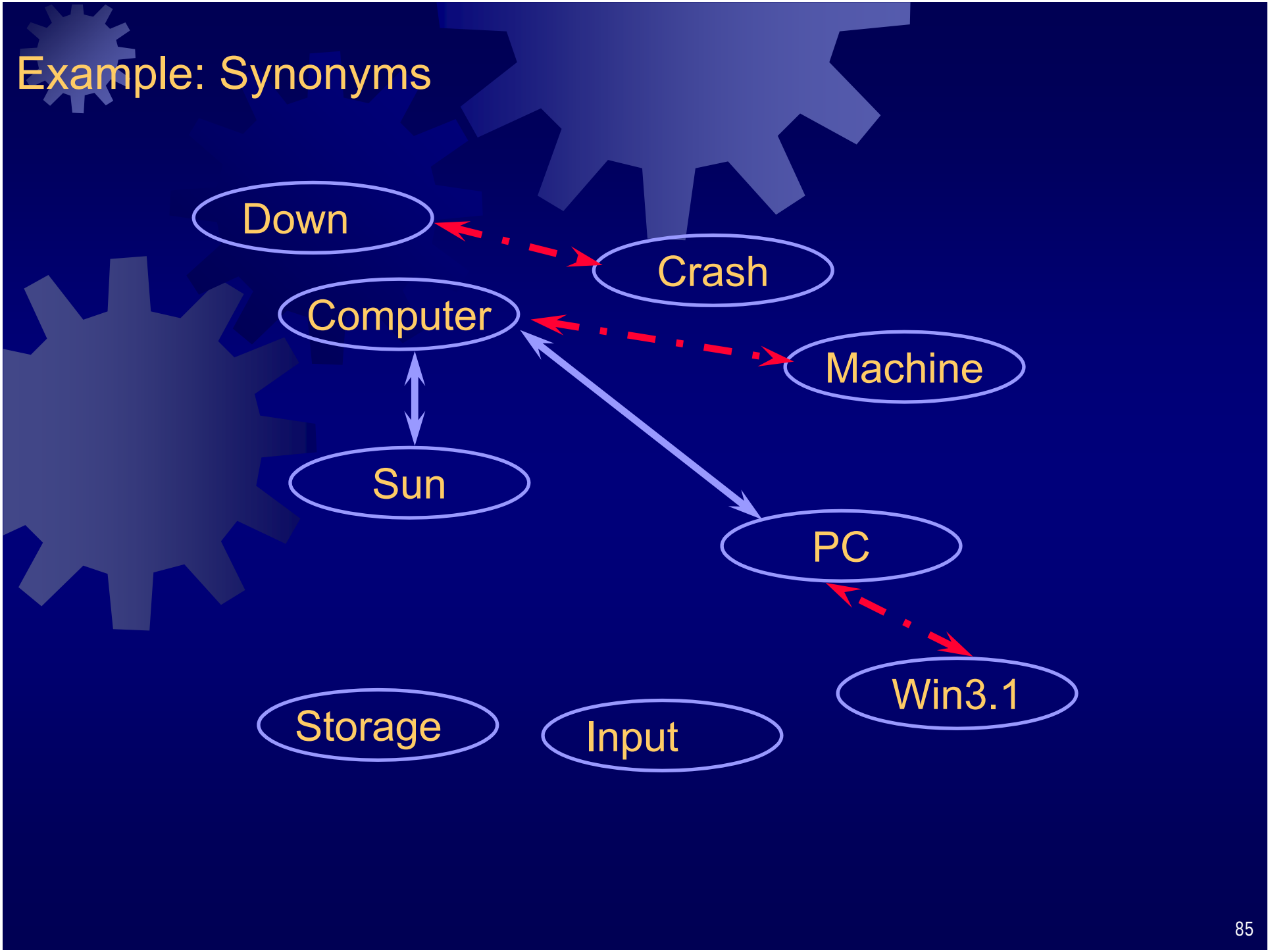
Example: Dictionary



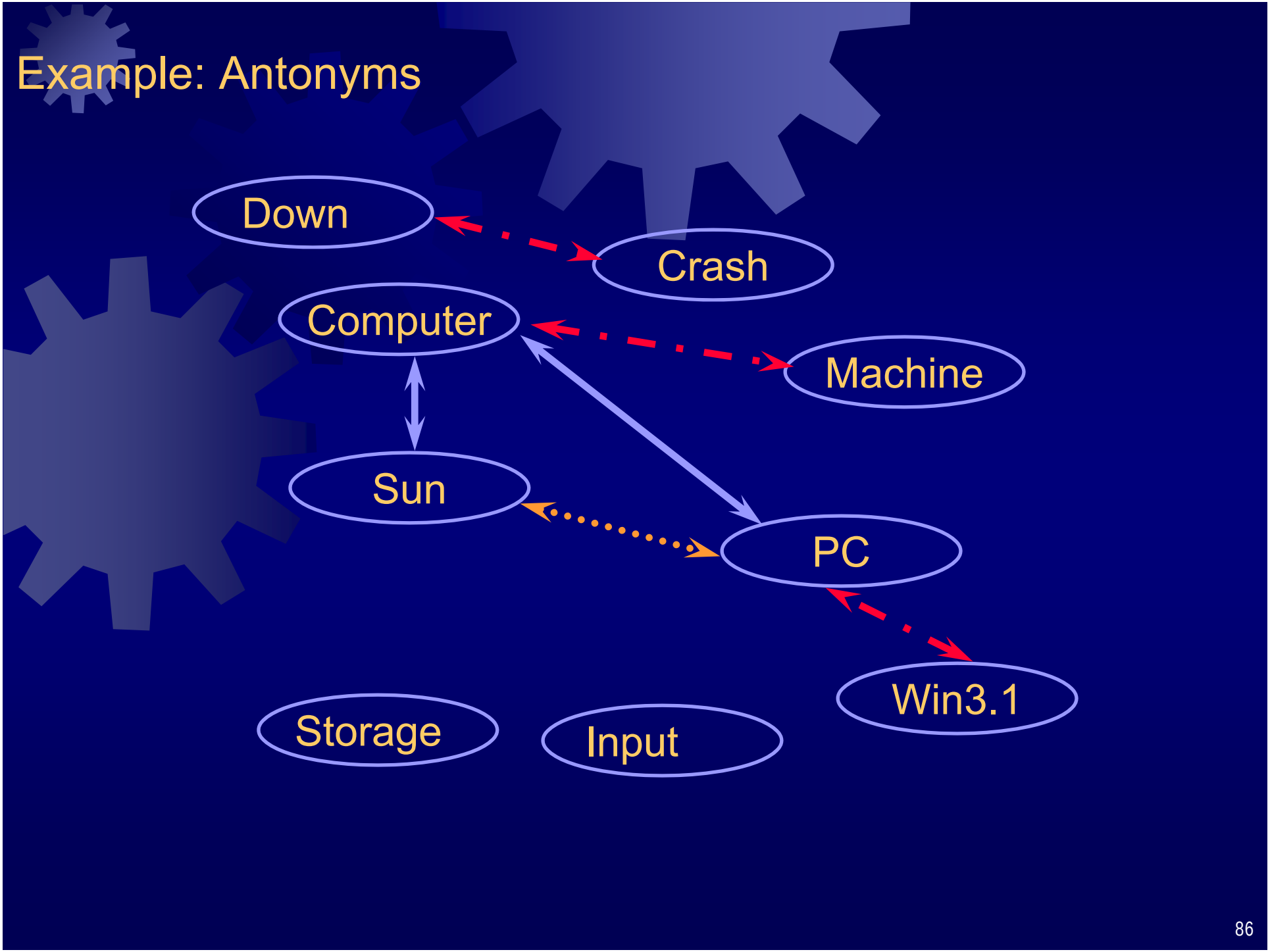
Example: Ontology



Example: Synonyms



Example: Antonyms



Example: Query

- **Q**: *On my PC the input of a long street name causes a crash. The error message is “Memoryfault”.*

Example: Query

- **Q**: *On my PC the input of a long street name causes a crash. The error message is "Memory fault".*

Example: Query and Results

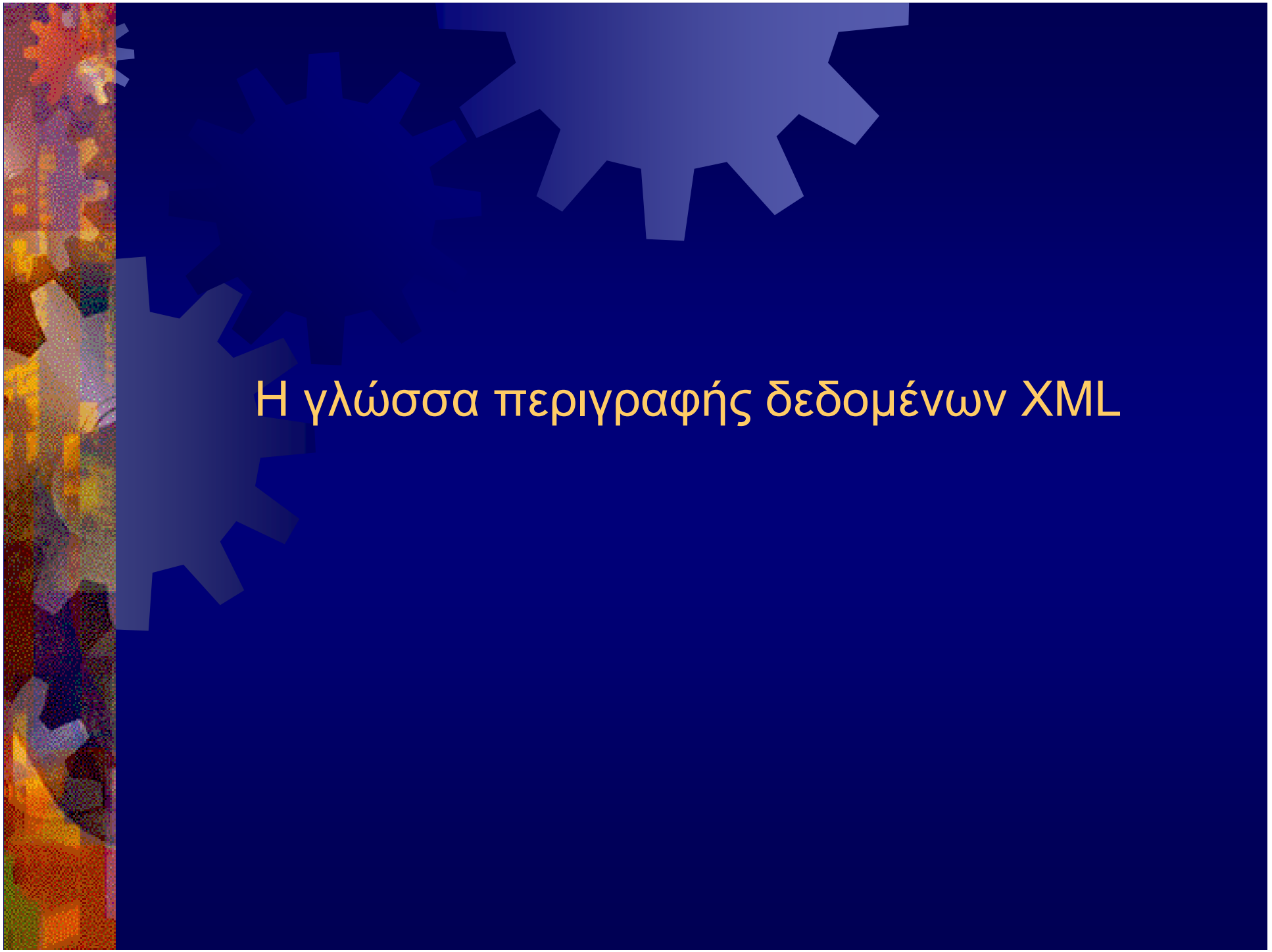
- **Q**: *On my PC the input of a long street name causes a crash. The error message is "Memory fault".*

- **F_1**: *On Windows 3.1 there is not enough memory allocated for the name of the street. This may cause the system to go down.*
- **F_2**: *The PC-Version stores the street name incorrectly.*
- **F_3**: *Typing German characters causes a Sun to crash.*

Example: Query and Results

- **Q**: *On my PC the input of a long street name causes a crash. The error message is "Memory fault".*

- **F_1**: *On Windows 3.1 there is not enough memory allocated for the name of the street. This may cause the system to go down.*
- **F_2**: *The PC-Version stores the street name incorrectly.*
- **F_3**: ~~*Typing German characters causes a Sun to crash.*~~


Η γλώσσα περιγραφής δεδομένων XML

XML - eXtensible Markup Language

- markup language σχετικά όμοια με την HTML
- σχεδιάστηκε να περιγράφει δεδομένα
- αποτελεί ένα εργαλείο μεταφοράς / μετάδοσης δεδομένων ανεξαρτήτως πλατφόρμας και λογισμικού
- Υποστηρίζει / διευκολύνει την ανταλλαγή δεδομένων μεταξύ μη συμβατών συστημάτων

Παράδειγμα σε XML

```
<book>
   <title>Nonmonotonic Reasoning: Context- Dependent
   Reasoning</title>
   <author>V. Marek</author>
   <author>M. Truszczynski</author>
   <publisher>Springer</publisher>
   <year>1993</year>
   <ISBN>0387976892</ISBN>
</book>
```

HTML versus XML: Ομοιότητες

- Πρώτον, και οι δύο αναπαραστάσεις χρησιμοποιούν 'ετικέτες' τα επονομαζόμενα *tags*, όπως <h2> και </year>
- Και οι δύο είναι *γλώσσες σήμανσης* (markup languages):
  - επιτρέπουν την εισαγωγή περιεχομένου και την παροχή πληροφοριών σχετικά με το ρόλο του περιεχομένου

HTML vs XML: Δομημένη πληροφορία

- Το πρόβλημα προκύπτει από το γεγονός ότι το κείμενο σε HTML δεν περιλαμβάνει πληροφορίες για τη δομή, δηλ. πληροφορίες για τμήματα του κειμένου και τις μεταξύ τους σχέσεις
- Σε αντίθεση το κείμενο σε XML είναι περισσότερο κατανοητό από τις μηχανές καθώς περιγράφεται κάθε τμήμα της πληροφορίας.
- Επιπλέον προσδιορίζονται οι *σχέσεις* μεταξύ των τμημάτων μέσω της χρήσης εμπερικλειόμενων tags.
  - Για παράδειγμα, τα tags <author> τοποθετούνται ανάμεσα στα tags <book> και επομένως περιγράφουν ιδιότητες του συγκεκριμένου βιβλίου.
- Μια μηχανή που επεξεργάζεται το συγκεκριμένο XML κείμενο συμπεράνει ότι
  - το στοιχείο 'author' αναφέρεται στο στοιχείο 'book' που το περιλαμβάνει
- XML επιτρέπει τον καθορισμό περιορισμών στις τιμές
  - πχ. ότι το έτος πρέπει να είναι τετραψήφιος αριθμός μικρότερος του 3000

Η XML

XML κείμενα αποτελούνται από:

- prolog
- αριθμό elements και attributes
- (προαιρετικά) epilog (δεν αναλύεται στη παρούσα)

Prolog

Το prolog αποτελείται από:

- Μία δήλωση XML και
- Μία προαιρετική αναφορά σε εξωτερικό αρχείο καθορισμού δομής κειμένων

<?xml version="1.0" encoding="UTF-16"?>

<!DOCTYPE book SYSTEM "book.dtd">

XML Elements

- Τα «πράγματα» στα οποία αναφέρεται το XML κείμενο
  - Π.χ. Βιβλία, συγγραφείς, εκδότες
- Ένα element αποτελείται από:
  - an opening tag
  - the content
  - a closing tag

<lecturer>David Billington</lecturer>

XML Elements (2)

- Υπάρχει σχεδόν πλήρης ελευθερία επιλογής Tag names
- Ο πρώτος χαρακτήρας πρέπει να είναι γράμμα, underscore, ή colon
- Κανένα όνομα δε πρέπει να ξεκινά με το λεκτικό “xml” ή συνδυασμούς αυτού
  - Π.χ. “Xml”, “xML”

Περιεχόμενο XML Elements

- Το περιεχόμενο μπορεί να είναι κείμενο, άλλα elements, ή τίποτα

```
<lecturer>
    <name>David Billington</name>
    <phone> +61 – 7 – 3875 507 </phone>
</lecturer>
```

- Στην XML κάποιοι χαρακτήρες είναι δεσμευμένοι και δεν μπορούν να τοποθετηθούν στο περιεχόμενο ενός κειμένου παρά μόνον με ειδικό τρόπο:
  - ο χαρακτήρας & μπορεί να εισαχθεί ως &
  - ο χαρακτήρας < μπορεί να εισαχθεί ως <
  - ο χαρακτήρας > μπορεί να εισαχθεί ως >
  - ο χαρακτήρας “ μπορεί να εισαχθεί ως "
  - ο χαρακτήρας ‘ μπορεί να εισαχθεί ως '

XML Attributes

- Το attribute είναι ένα ζεύγος name-value pair μέσα στο opening tag ενός element

```
<lecturer name="David Billington" phone="+61 – 7 – 3875
507"/>
```

Παράδειγμα XML Attribute

```
<order orderNo="23456" customer="John Smith"
       date="October 15, 2002">
      <item itemNo="a528" quantity="1"/>
      <item itemNo="c817" quantity="3"/>
</order>
```

Το ίδιο παράδειγμα χωρίς Attributes

```
<order>
   <orderNo>23456</orderNo>
   <customer>John Smith</customer>
   <date>October 15, 2002</date>
   <item>
        <itemNo>a528</itemNo>
        <quantity>1</quantity>
   </item>
   <item>
        <itemNo>c817</itemNo>
        <quantity>3</quantity>
        </item>
</order>
```

XML Elements vs Attributes

- Τα Attributes μπορούν να υποκατασταθούν από elements
- Πότε χρησιμοποιεί κανείς attributes ή elements είναι θέμα επιλογής
- Υπάρχουν οι εξής περιορισμοί
  - Εαν υπάρχει περίπτωση να έχουμε δύο ή περισσότερες τιμές για το συγκεκριμένο είδος πληροφορίας, χρησιμοποιούμε element

```
<book>
    <author>John Mavropoulos</author>
    <author>Jim Pavlopoulos</author>
</book>
```

  - Τα attributes **δεν μπορούν** να είναι nested

Δόμηση XML κειμένων

- είναι απαραίτητο να οριστούν τα ονόματα όλων των στοιχείων (elements) και των χαρακτηριστικών (attributes) που θα χρησιμοποιηθούν
- πρέπει να καθοριστεί η δομή: τι τιμές μπορεί να πάρει κάποιο χαρακτηριστικό, ποια στοιχεία μπορούν ή πρέπει να περιλαμβάνονται μέσα σε άλλα στοιχεία, κλπ.
- ένα XML κείμενο είναι έγκυρο (valid) εάν είναι καλά μορφοποιημένο και περιλαμβάνει κανόνες δόμησης τους οποίους και ακολουθεί
- Υπάρχουν δύο τρόποι ορισμού της δομής των XML κειμένων:
  - τα DTDs, ο παλαιότερος τρόπος με τους περισσότερους περιορισμούς
  - το XML Schema, το οποίο προσφέρει πολλαπλές δυνατότητες κυρίως για τον ορισμό των τύπων δεδομένων (data types)

DTD: Element Type Definition

```
<lecturer>
    <name>David Billington</name>
    <phone> +61 - 7 - 3875 507 </phone>
</lecturer>
```

DTD for above element (and all **lecturer** elements):

```
<!ELEMENT lecturer (name,phone)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT phone (#PCDATA)>
```

Η σημασία του DTD

- Οι τύποι στοιχείων lecturer, name και phone μπορούν να χρησιμοποιηθούν στο κείμενο
- Το στοιχείο lecturer περιλαμβάνει το στοιχείο name και το στοιχείο phone , με αυτή τη σειρά
- Ένα στοιχείο name και ένα στοιχείο phone μπορούν να έχουν οποιοδήποτε περιεχόμενο

Χρήση σχημάτων

- Ένα DTD μπορεί να είναι ορισμένο μέσα στο XML (inline) ή έξω από αυτό (reference)
- Internal DTD Declaration
  - <!DOCTYPE root-element [element-declarations]>

Παράδειγμα inline χρήσης

- Χρήση internal DTD:

```
<?xml version="1.0"?>
<!DOCTYPE note [ <!ELEMENT note (to,from,heading,body)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
<!ELEMENT heading (#PCDATA)>
<!ELEMENT body (#PCDATA)> ]>
<note>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>Don't forget me this weekend</body>
</note>
```

Χρήση σχημάτων

- External DTD Declaration
- <!DOCTYPE root-element SYSTEM **"filename"**>

Παράδειγμα external χρήσης

```
<?xml version="1.0"?>
<!DOCTYPE note SYSTEM "note.dtd">
<note>
    <to>Tove</to>
    <from>Jani</from>
    <heading>Reminder</heading>
    <body>Don't forget me this weekend!</body>
</note>
```

Ορισμός PCDATA

- PCDATA = parsed character data

- Ορισμός w3c: “PCDATA is text that WILL be parsed by a parser. The text will be examined by the parser for entities and markup. Tags inside the text will be treated as markup and entities will be expanded. ”

- Τα PCDATA ΔΕΝ επιτρέπεται να περιέχουν τους χαρακήρες &, <, or >
- Αντί αυτών γίνεται χρήση των & < and &gt

DTD: Περιεχόμενο elements

- Elements με Parsed Character Data
  - <!ELEMENT element-name (#PCDATA)>
  - Π.χ.:<!ELEMENT from (#PCDATA)>
- Elements με παιδιά
  - <!ELEMENT element-name (child1)> ή
  - <!ELEMENT element-name (child1,child2,...)>
  - π.χ.:<!ELEMENT note (to,from,heading,body)>

Πλήθος εμφάνισης (occurrence)

- Μία εμφάνιση:
  - <!ELEMENT note (message)>
  - No cardinality operator means exactly once
- Μία η περισσότερες
  - <!ELEMENT note (message+)>
- Καμία η περισσότερες
  - <!ELEMENT note (message*)>
- Μία ή καμία
  - <!ELEMENT note (message?)>
- Ή το ένα Ή το άλλο
  - <!ELEMENT note (to,from,header,(message|body))>

DTD: Attributes

- Στο DTD, τα attributes δηλώνονται με ATTLIST
- <!ATTLIST element-name attribute-name attribute-type default-value>

DTD: Attribute Types

- Similar to predefined data types, but limited selection
- The most important types are
  - **CDATA**, string (sequence of characters)
    - Ορισμός w3c: "**CDATA is text that will NOT be parsed by a parser**. Tags inside the text will NOT be treated as markup and entities will not be expanded."
  - **ID**, τιμή που είναι μοναδική σε όλο το XML κείμενο
  - **IDREF**, μία αναφορά στο ID ενός άλλου element
  - **IDREFS**, μια λίστα από μία ή περισσότερες αναφορές σε IDs
  - **(v1| . . . |vn)**, μια λίστα τιμών οι οποίες χωρίζονται από κάθετες γραμμές. Το attribute παίρνει μία από αυτές τις τιμές
- Limitations: no dates, number ranges etc.

DTD: Default Values

- **#REQUIRED**
  - Attribute must appear in every occurrence of the element type in the XML document
- **#IMPLIED**
  - The appearance of the attribute is optional
- **#FIXED "value"**
  - Every element must have this attribute
- **#DEFAULT "value"**
  - This specifies the default value for the attribute

Παράδειγμα

```
<order orderNo="23456"
                customer="John Smith"
                date="October 15, 2002">
      <item itemNo="a528" quantity="1"/>
      <item itemNo="c817" quantity="3"/>
</order>
```

DTD for the above elements and attributes:

```
<!ELEMENT order (item+)>
<!ATTLIST order orderNo      ID          #REQUIRED
                customer     CDATA       #REQUIRED
                date         CDATA       #REQUIRED>

<!ELEMENT item EMPTY>
<!ATTLIST item  itemNo       ID          #REQUIRED
                quantity     CDATA       #REQUIRED
                comments     CDATA       #IMPLIED>
```

XML Schema ή XML Schema Definition (XSD)

- παρέχει μια ιδιαίτερα πλούσια γλώσσα για τον ορισμό της δομής ενός XML κειμένου
- η σύνταξή που χρησιμοποιεί βασίζεται στην ίδια την XML
  - βελτιώνει την αναγνωσιμότητα
  - επιτρέπει την ουσιαστική επαναχρησιμοποίηση της τεχνολογίας
    - Δεν υπάρχει πλέον η ανάγκη να δημιουργηθούν ξεχωριστοί συντακτικοί αναλυτές (parsers), κειμενογράφοι (editors), κλπ.
- πιθανότητα επαναχρησιμοποίησης και επαναπροσδιορισμού των XML Schemas
  - Ένα XML Schema επιτρέπει τον ορισμό νέων τύπων με την επέκταση ή τον περιορισμό των ήδη υπαρχόντων
- παρέχει ένα πολυπληθές σύνολο τύπων δεδομένων (datatypes) που μπορούν να χρησιμοποιηθούν στα XML κείμενα
  - τα DTDs περιλαμβάνουν μόνο στοιχειοσειρές (strings)

XML Schema (2)

- Ένα XML Schema είναι ένα στοιχείο με tag έναρξης:
- <xsd:schema xmlnsQxsd="http://www.w3.org/2000/10/XMLSchema" version="1.0">
- Το στοιχείο χρησιμοποιεί το XML Schema που βρίσκεται στο δικτυακό χώρο του W3C
  - είναι η βάση στην οποία μπορούν να στηριχτούν τα νέα schemas

Element Types

```
<element name="email"/>
<element name="head" minOccurs="1"      maxOccurs="1"/>
<element name="to" minOccurs="1"/>
```

- Cardinality constraints:
- minOccurs="x" (default value 1)
- maxOccurs="x" (default value 1)
- Generalizations of *,?,+ offered by DTDs

Attribute Types

```
<attribute name="id" type="ID" use="required"/>
< attribute name="speaks" type="Language"
   use="default" value="en"/>
```

- Existence: use="x", where x may be optional or required
- Default value: use="x" value="...", where x may be default or fixed

Data Types

- There is a variety of built-in data types
  - Numerical data types: integer, Short etc.
  - String types: string, ID, IDREF, CDATA etc.
  - Date and time data types: time, Month etc.
- There are also user-defined data types
  - simple data types, which cannot use elements or attributes
  - complex data types, which can use these

Data Types (2)

- Complex data types are defined from already existing data types by defining some attributes (if any) and using:
  - sequence, a sequence of existing data type elements (order is important)
  - all, a collection of elements that must appear (order is not important)
  - choice, a collection of elements, of which one will be chosen

A Data Type Example

```
<complexType name="lecturerType">
    <sequence>
        <element name="firstname" type="string"
            minOccurs="0“   maxOccurs="unbounded"/>
        <element name="lastname" type="string"/>
    </sequence>
    <attribute name="title" type="string"
        use="optional"/>
</complexType>
```

Data Type Extension

- Already existing data types can be extended by new elements or attributes. Example:

```
<complexType name="extendedLecturerType">
    <extension base="lecturerType">
       <sequence>
             <element name="email" type="string"
                   minOccurs="0" maxOccurs="1"/>
       </sequence>
       <attribute name="rank" type="string"
    use="required"/>
    </extension>
</complexType>
```

Data Type Restriction

- An existing data type may be restricted by adding constraints on certain values

Restriction of Simple Data Types

```
<simpleType name="dayOfMonth">
      <restriction base="integer">
            <minInclusive value="1"/>
            <maxInclusive value="31"/>
      </restriction>
</simpleType>
```

Data Type Restriction: Enumeration

```
<simpleType name="dayOfWeek">
      <restriction base="string">
            <enumeration value="Mon"/>
            <enumeration value="Tue"/>
            <enumeration value="Wed"/>
            <enumeration value="Thu"/>
            <enumeration value="Fri"/>
            <enumeration value="Sat"/>
            <enumeration value="Sun"/>
      </restriction>
</simpleType>
```

Namespaces

- Ένα xml κείμενο μπορεί να χρησιμοποιεί παραπάνω από ένα DTD ή XML Schema
  - Καθώς κάθε κείμενο δόμησης (DTD ή XML Schema) μπορεί να αναπτύχθηκε ανεξάρτητα, ενδέχεται να υπάρχουν συγκρούσεις ονομάτων, δηλ. ίδια ονόματα να χρησιμοποιήθηκαν με διαφορετικό σκοπό σε διαφορετικά κείμενα δόμησης
  - Το πρόβλημα αυτό ξεπερνιέται με τη χρήση διαφορετικού προθέματος (prefix:name) για κάθε κείμενο δόμησης
  - Μία δήλωση namespace έχει την μορφή:
    - xmlns:prefix="location"
  - όπου location είναι η διεύθυνση του DTD ή XML schema.
- If a prefix is not specified: xmlns="location" then the location is used by default
- Namespaces are declared within an element and can be used in that element and any of its children (elements and attributes)

Παράδειγμα

```
<vu:instructors   xmlns:vu="http://www.vu.com/empDTD"
                  xmlns:gu="http://www.gu.au/empDTD"
                  xmlns:uky="http://www.uky.edu/empDTD">
    <uky:faculty  uky:title="assistant professor"
                  uky:name="John Smith"
                  uky:department="Computer Science"/>
    <gu:academicStaff   gu:title="lecturer"
                        gu:name="Mate Jones"
                        gu:school="Information Technology"/>
</vu:instructors>
```