



ΑΛΛΗΛΕΠΙΔΡΑΣΗ ΑΝΘΡΩΠΟΥ-ΥΠΟΛΟΓΙΣΤΗ

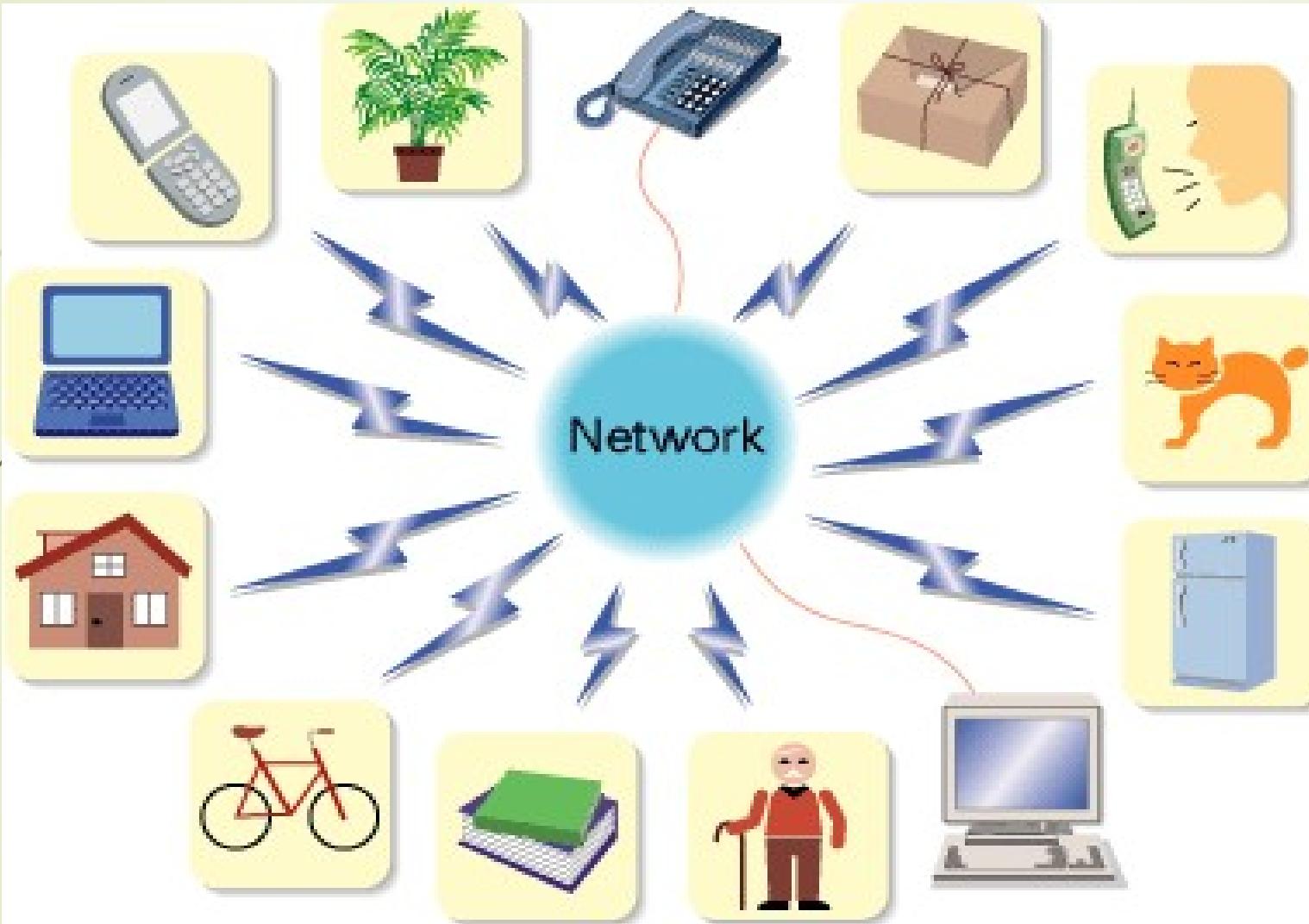
ΜΑΘΗΜΑ 5^ο

Μοντελοποίηση Χρηστών – Recommendation Συστήματα

Παναγιώτης Τσανάκας
Γιώργος Σιόλας
Μάριος Κόνιαρης
Γιώργος Αλεξανδρίδης

7ο ΕΞΑΜΗΝΟ ΣΧΟΛΗΣ ΗΜΜΥ, ΡΟΗ Υ

1.1. Φορητότητα & διάχυτη υπολογιστικότητα - Η υπό εξαφάνιση διεπιφάνεια χρήσης

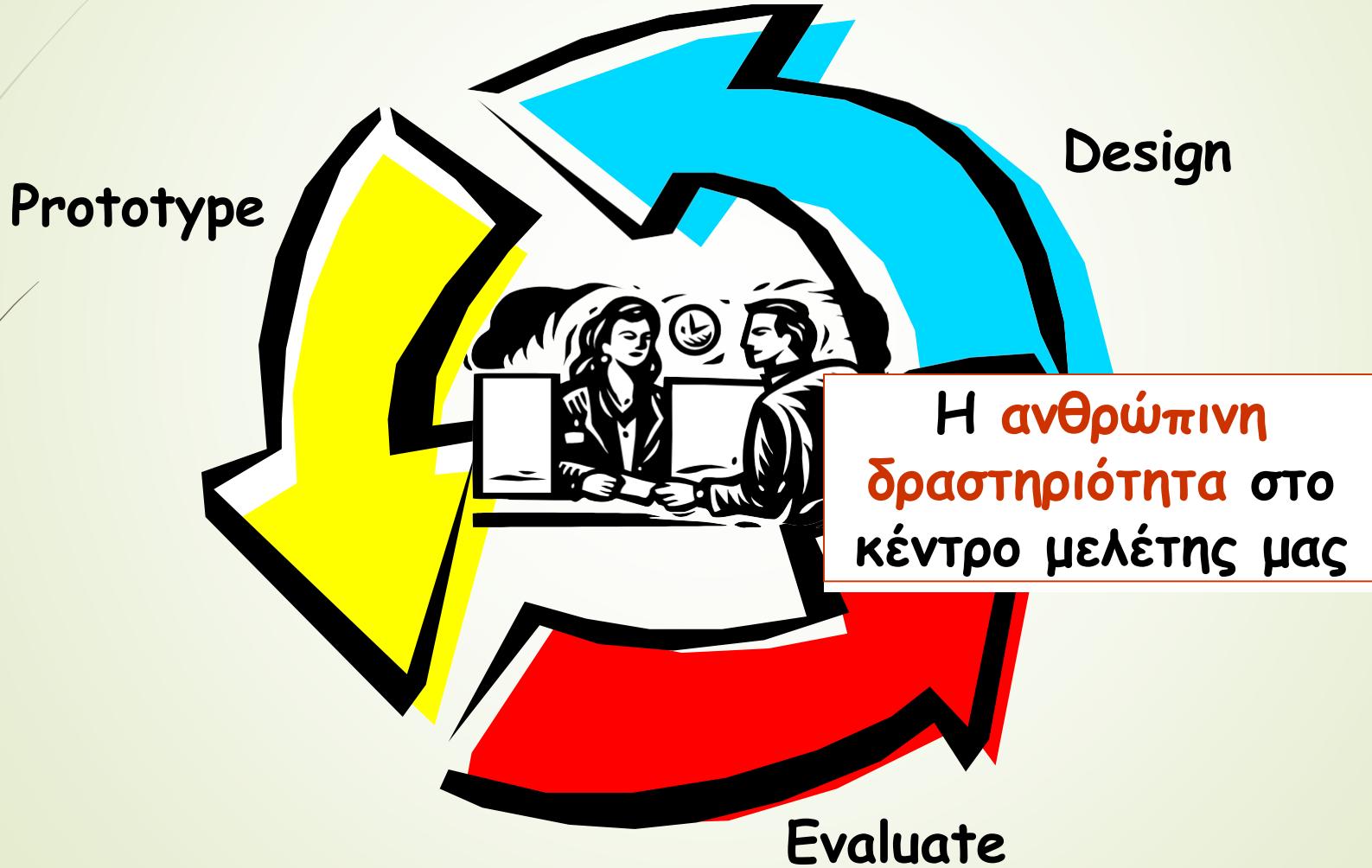


Ιθικούτων συμμετίνα will enable diverse wireless applications

1.2 Ευχρηστία (usability)

- ▶ Η δυνατότητα ενός προϊόντος που χρησιμοποιείται από
- ▶ **καθορισμένους χρήστες με καθορισμένους στόχους,**
- ▶ **υπό καθορισμένες συνθήκες χρήσης**
- ▶ Και να είναι:
 - ▶ αποτελεσματικό (**effectiveness**),
 - ▶ αποδοτικό (**efficiency**)
 - ▶ και να παρέχει υποκειμενική ικανοποίηση (**satisfaction**) στους χρήστες του

1.3 Ανθρωπο-κεντρική σχεδίαση



2. ΑΝΑΛΥΣΗ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΩΝ ΧΡΗΣΤΩΝ

- Προφίλ Κάθε Χρήστη
- Κατηγορίες Χρηστών
- Μοντελοποίηση Χρηστών
- Στατικά – Προσαρμοζόμενα Μοντέλα
- Πλαίσιο Αλληλεπίδρασης
- Ανάλυση Εργασιών vs Κατηγοριών Χρηστών

2.1 Ανάλυση Ενδιαφερομένων/Χρηστών Κατηγορίες Stakeholders

- (1) οι εμπλεκόμενοι στη **σχεδίαση και ανάπτυξη** του συστήματος, το τεχνικό προσωπικό, προγραμματιστές, αναλυτές, συγγραφείς εγχειριδίων.
- (2) οι υπεύθυνοι για την **εισαγωγή, εγκατάσταση και συντήρηση** του, όπως οι υπεύθυνοι συντήρησης του εξοπλισμού, εκπαίδευσης του προσωπικού.
- (3) οι έχοντες **οικονομικό συμφέρον** συνδεδεμένο με την κατασκευή ή αγορά του συστήματος, όπως ο πάροχος του προϊόντος ή ο αγοραστής του.
- (4) οι ενδιαφερόμενοι για την **χρήση** του, όπως η διοίκηση μιας επιχείρησης καθώς και οι χρήστες του.

2.2 Χαρακτηριστικά Χρηστών-Εργασιών/1

- ▶ (Α) Χαρακτηριστικά των χρηστών :
- ▶ **Ατομικά χαρακτηριστικά:** Ηλικία, φυσικές ικανότητες/ ιδιαιτερότητες, μαθησιακή ικανότητα, γνωσιακή ικανότητα, εμπειρία-δεξιότητες, κίνητρα και φιλοδοξίες, πολιτισμικό υπόβαθρο, φοβίες, προσωπικότητα.
- ▶ **Χαρακτηριστικά σε σχέση με υπολογιστές:** προηγούμενη εμπειρία σε λογισμικό και λειτουργικά συστήματα, εμπειρία σε χρήση συσκευών, προδιάθεση έναντι υπολογιστών και πληροφορικής.
- ▶ **Ομαδικά χαρακτηριστικά:** Στόχοι και αποστολή ομάδας, συνοχή και ομοιογένεια μελών ομάδας, αυτονομία, εξάρτηση από άλλες ομάδες, δομή και δυναμικά χαρακτηριστικά, κύρος, αυτόβουλη ή καταναγκαστική συμμετοχή.

Χαρακτηριστικά Χρηστών-Εργασιών/2

(Β) Περιβάλλον :

Περιβάλλον εργασίας : Συνθήκες θορύβου, κρύου, υγρασίας, σκόνη καθαριότητα, χρήση επικίνδυνων τοξικών ουσιών

Οργάνωση χώρου εργασίας : Κανάλια επικοινωνίας χρηστών, οργανωτική δομή, επίδραση αυτοματισμού στην πρακτική εργασίας και περιεχόμενο εργασίας, καταγραφή πιθανών απωλειών ικανοτήτων, απώλεια απασχόλησης, μεταβολές στη δομή εξουσίας, αποκέντρωση/ συγκέντρωση εξουσίας.

(Γ) Συνθήκες απασχόλησης των χρηστών. Φύλλα περιγραφής μιας τυπικής ημέρας στη ζωή του εργαζόμενου με σχόλια για το πώς αυτή θα μεταβληθεί όταν εισαχθεί το προτεινόμενο σύστημα.

Χαρακτηριστικά Χρηστών-Εργασιών/3

(Δ) Ανάλυση βασικών εργασιών

Οργανωτική ανάλυση: σπουδαιότητα εργασίας, θέματα ασφάλειας, κίνητρα εκτέλεσής της, απαιτούμενο επίπεδο ικανότητας για την εκπόνησή της, εξάρτηση από άλλες εργασίες,

Ανάλυση χρονικών περιορισμών: συχνότητα εκτέλεσης της εργασίας, μέσος απαιτούμενος χρόνος περάτωσής της, χρόνος προετοιμασίας, τμηματοποίηση, δυνατότητα εκτέλεσής της κατά διακριτά τμήματα

Ανάλυση ανθρώπινης εμπλοκής : υποχρεωτικός/ προαιρετικός χαρακτήρας, προκαλούμενη πίεση, κριτήρια απόδοσης.

3. ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ (RECOMMENDER)

- **Τα δομικά συστατικά**

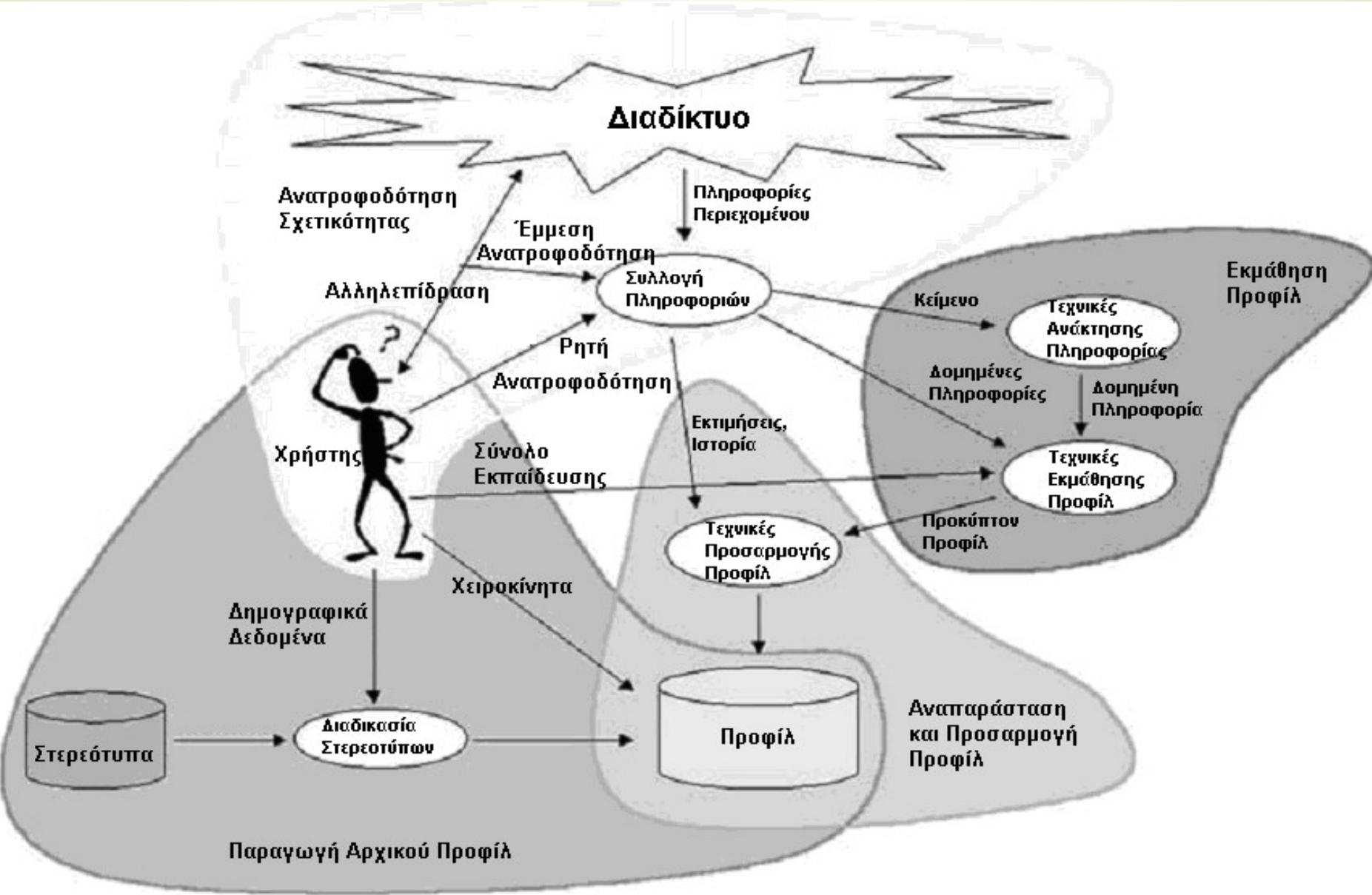
ΠΑΡΑΓΩΓΗ ΚΑΙ ΣΥΝΤΗΡΗΣΗ ΠΡΟΦΙΛ

- Αναπαράσταση Προφίλ
- Παραγωγή αρχικού προφίλ
- Τεχνικές εκμάθησης προφίλ
- Ανατροφοδότηση σχετικότητας
- Τεχνικές προσαρμογής προφίλ

ΕΚΜΕΤΑΛΛΕΥΣΗ ΠΡΟΦΙΛ

- Μέθοδοι φιλτραρίσματος πληροφοριών
- Ταίριασμα στοιχείων- προφίλ χρήστη
- Ταίριασμα παραμέτρων χρήστη

Παραγωγή και Συντήρηση Προφίλ



3.1 Αναπαράσταση Προφίλ

- **Μοντέλα βασισμένα στην Ιστορία**
(ηλεκτρονικό εμπόριο, Amazon.com και CDNow.com, ηλεκτρονικό ταχυδρομείο Tapestry)
- **Μοντέλα βασισμένα σε Δημογραφικά χαρακτηριστικά γνωρίσματα** (μέσω Στερεοτύπων)
- **Μοντέλα Διανυσματικού Χώρου**
(διάνυσμα χαρακτηριστικών γνωρισμάτων, συνήθως λέξεων ή εννοιών, με μια σχετική τιμή, Boolean ή πραγματικό αριθμό: κάθε έγγραφο **διάνυσμα -λέξεων** - σε ένα διανυσματικό χώρο έτσι ώστε έγγραφα με **παρόμοιο περιεχόμενο** να έχουν παρόμοια διανύσματα. Το ίδιο και για τους χρήστες.)

3.1 Αναπαράσταση Προφίλ

- **Μοντέλα βασισμένα σε Ταξινομητές**

Τα συστήματα που χρησιμοποιούν ταξινομητή ως τεχνική εκμάθησης προφίλ χρήστη διατηρούν τη δομή του:

- Νευρωνικά δίκτυα, κατανεμημένες αρχιτεκτονικές που μαθαίνουν συσχετίσεις των δεδομένων από παραδείγματα
- Δένδρα αποφάσεων, σύνολα κόμβων και κατευθυνόμενων ακμών που συνδέουν τους κόμβους (δομή δέντρων)
- Κανόνες (association /inducted rules), π.χ. μια ορισμένη ταινία αγοράζεται συχνά μαζί με άλλες
- Μπεϋζιανά δίκτυα, κατευθυνόμενοι ακυκλικοί γράφοι (κόμβοι → προτασιακές μεταβλητές, ακμές → εξαρτήσεις. Η αξία του κόμβου είναι συνάρτηση των κόμβων από όπου εξαρτάται)

3.2 Παραγωγή αρχικού προφίλ

- **Κενό**
- **Χειροκίνητο** (χρήστες καταχωρούν τα ενδιαφέροντά τους: λέξεις κλειδιά, θέματα →κουραστικό, ασαφές)
- **Στερεότυπο** (όνομα, διεύθυνση, περιοχή, πόλη, ηλικία, φύλο, τρόπος ζωής →προσωπικά στοιχεία!)
- **Σύνολο εκπαίδευσης** (ζητείται από το χρήστη να εκτιμήσει μερικά συγκεκριμένα παραδείγματα , σχετικά ή άσχετα με τα ενδιαφέροντά του →απλό, αλλά μπορεί να εισάγει λάθη)

3.3 Τεχνικές εκμάθησης προφίλ

- **Μη απαραίτητη** (Συστήματα παίρνουν πληροφορίες παραμέτρων χρήστη από βάση δεδομένων, π.χ., προϊόντων \bowtie προφίλ = κατάλογος αγορών, ή δεν αλλάζουν το αρχικό - μέσω στερεοτύπων – προφίλ)
- **Τεχνικές ανάκτησης δομημένων πληροφοριών**
(κείμενο: α) απαλοιφή λέξεων τερματισμού, άρθρων, θεώρηση του θέματος λέξεων β) ευρετηρίαση των πληροφοριών: συχνότητα εμφάνισης λέξεων \bowtie πιθανή σχετικότητα ενός στοιχείου, TF/IDF μέθοδος: Οι όροι που εμφανίζονται συχνά σε ένα έγγραφο (TF = συχνότητα όρου), αλλά σπάνια εκτός εγγράφου (IDF = αντίστροφη συχνότητα εγγράφου), είναι πιθανότερο να είναι σχετικοί με το θέμα του εγγράφου.

3.3 Τεχνικές εκμάθησης προφίλ

- **Συσταδοποίηση (Clustering)**, ομαδοποίηση χρηστών με παρόμοια χαρακτηριστικά, ταίριασμα χρήστη με συστάδες χρηστών παρόμοιου προφίλ
- **Ταξινομητές**: τεχνικές μηχανικής μάθησης → νευρωνικά δίκτυα, δέντρα απόφασης (ID3), κανόνες & επαγωγική μάθηση (C4.5), Μπεϋζιανοί αλγόριθμοι

3.4 Ανατροφοδότηση σχετικότητας

- Τα ανθρώπινα ενδιαφέροντα αλλάζουν με τον χρόνο.
- Θετικές και αρνητικές πληροφορίες.
- **Καμία ανατροφοδότηση** (μόνο χειροκίνητες αλλαγές προφίλ)
- **Ρητή ανατροφοδότηση** (πόσο σχετικό ή ενδιαφέρον είναι ένα στοιχείο για το χρήστη, ή πόσο νομίζει πως είναι το στοιχείο για άλλους χρήστες, μόνο 15% χρηστών απαντάει ειδικά σε αρνητική σχετικότητα)
- **Εμμεση ανατροφοδότηση** (ανάλυση συμπεριφοράς, π.χ. διάσωση, εκτύπωση, διαγραφή εγγράφου, προσθήκη στα αγαπημένα, απάντηση ή προώθηση μηνύματος, μεγιστοποίηση/ελαχιστοποίηση/επαναφορά του παραθύρου που περιέχει το έγγραφο ή την ιστοσελίδα)
- **Υβριδική προσέγγιση** (έμμεση για μείωση προσπαθειών χρήστη, ρητή για ακρίβεια απόκρισης)

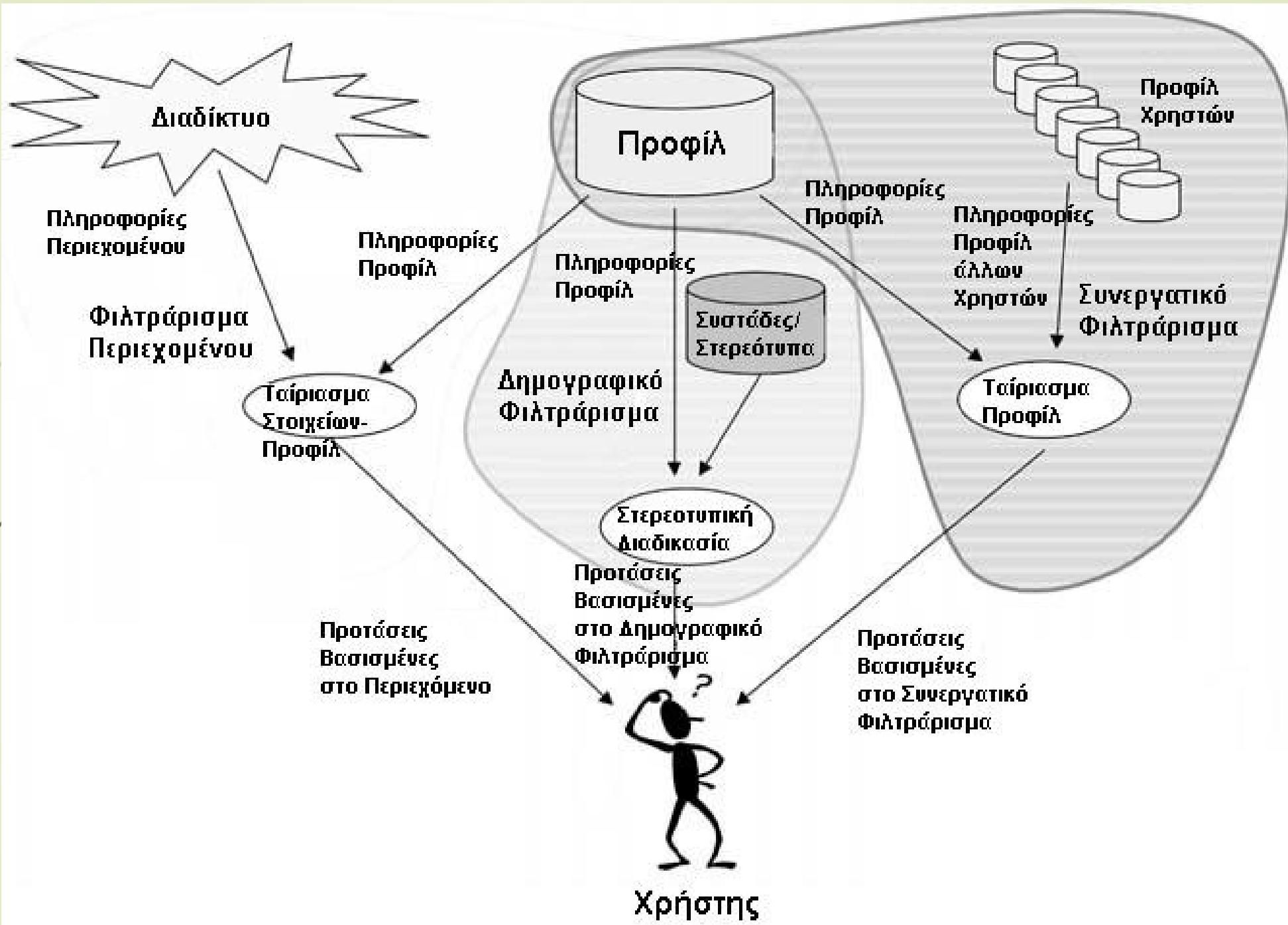
3.5 Τεχνικές προσαρμογής προφίλ

- **Χειροκίνητη** (δύσκολη)
- **Προσθήκη νέων πληροφοριών** (δεν ξεχνά όμως τις παλιές πληροφορίες)
- **Συνάρτηση βαθμιαίας λήθης** (εισάγοντας χρονικά παράθυρα)
- **Φυσική επιλογή** (γενετικοί αλγόριθμοι, με σύστημα πρακτόρων: οι πράκτορες που παράγουν τα καλύτερα αποτελέσματα αναπαράγονται με τεχνικές διασταυρώσεων και μεταλλαγής και οι άλλοι καταστρέφονται)

4. ΕΚΜΕΤΑΛΛΕΥΣΗ ΠΡΟΦΙΛ ΧΡΗΣΤΩΝ

Τρεις κύριες διαστάσεις χαρακτηρίζουν τους πράκτορες των ευφυών recommender συστημάτων:

- η μέθοδος **φιλτραρίσματος** πληροφοριών (δημογραφική, περιεχομένου και συνεργατική → υβριδική),
- το **ταίριασμα δεδομένων** με το προφίλ χρηστών (περιεχομένου)
- οι τεχνικές **ταιριάσματος προφίλ** (συνεργατικές).



4.1 Μέθοδοι φίλτραρισμάτος πληροφοριών

Δημογραφικό φίλτραρισμα

- Τα προφίλ χρηστών δημιουργούνται με ταξινόμηση των χρηστών σε στερεοτυπικές περιγραφές → χαρακτηριστικά γνωρίσματα κατηγοριών χρηστών.
- Τα προσωπικά στοιχεία του χρήστη είναι απαραίτητα και χρησιμοποιούνται για την ταξινόμηση. Συνήθως λαμβάνονται κατά την αίτηση εγγραφής στο σύστημα.
- Τα προφίλ που προκύπτουν επεκτείνουν το εύρος πληροφοριών της δημογραφικής βάσης δεδομένων.

Γενικές Συστάσεις – Οχι προσαρμογή.



Φιλτράρισμα βασισμένο στο Περιεχόμενο

- Συστήνουν δεδομένα επειδή είναι παρόμοια με τα δεδομένα για τα οποία ο χρήστης έχει εκδηλώσει ενδιαφέρον στο παρελθόν (π.χ. ενδιαφέρον ή όχι στοιχείο).
- Τα προφίλ χρηστών δημιουργούνται χρησιμοποιώντας τα χαρακτηριστικά γνωρίσματα που εξάγονται από αυτά τα δεδομένα.
- Κάθε χρήστης υποτίθεται ότι λειτούργησε ανεξάρτητα.

Επιφανειακή Συσχέτιση περιεχομένου

Συνεργατικό φίλτραρισμα

- Στατιστική ανάλυση των αποκρίσεων που δίνονται **ρητά** από διαφορετικούς χρήστες ή **σιωπηρά** με την παρακολούθηση της συμπεριφοράς τους.
- Αντί να υπολογίζεται η ομοιότητα των πληροφοριών/εγγράφων, υπολογίζεται η **ομοιότητα μεταξύ των χρηστών**.
- Το προφίλ χρήστη αποτελείται από τα στοιχεία που ο χρήστης έχει διευκρινίσει. Αυτά συγκρίνονται με εκείνα άλλων χρηστών.
- Για κάθε χρήστη καθορίζεται ένα σύνολο "**πλησιέστερων γειτόνων**" χρησιμοποιώντας το συσχετισμό μεταξύ των δεδομένων που έχουν συλλεγεί για αυτούς. Η τεχνική **καλύτερων N-γειτόνων** είναι η επιλογή ενός αριθμού N χρηστών που ταιριάζουν περισσότερο με τον συγκεκριμένο.
- Τα μη ρητά εκπεφρασμένα δεδομένα προβλέπονται χρησιμοποιώντας έναν συνδυασμό των αποτελεσμάτων από τους πλησιέστερους γείτονες.

Ανεπαρκές αν υπάρχουν λίγοι χρήστες, χρήστες με ειδικά ενδιαφέροντα, δύσκολη προσαρμογή αν πρέπει να γίνεται σε όλους τους σχετικούς χρήστες.

4.2 Ταίριασμα δεδομένων- προφίλ χρήστη

Αμεση σύγκριση μεταξύ των παραμέτρων χρήστη και των πληροφοριών/εγγράφων

Τεχνικές ταιριάσματος προφίλ-δεδομένων

- **Ομοιότητα συνημίτονου** (αναπαράσταση χρήστη και εγγράφου ως ένα διάνυσμα με στοιχεία: το βαθμό συσχέτισης με το σύνολο χαρακτηριστικών ή τις κατηγορίες ενδιαφερόντων)
- **Πλησιέστερος γείτονας** (π.χ. Case-Based Reasoning, όπου το προφίλ ενός χρήστη αναπαρίσταται από μια συλλογή των προηγούμενων εμπειριών του)
- **Ταξινόμηση** (νευρωνικά δίκτυα, δέντρα αποφάσεων)

5. ΠΑΡΑΔΕΙΓΜΑΤΑ ΣΥΣΤΗΜΑΤΩΝ

Συστήματα recommender ιστοσελίδων.

- Συνήθως τα προφίλ των χρηστών αναπαρίστανται με διανύσματα χαρακτηριστικών γνωρισμάτων.
- Η εκμάθηση των προφίλ γίνεται από το κείμενο μέσω της επιλογής χαρακτηριστικών γνωρισμάτων και των τεχνικών TF-IDF.
- Το ταίριασμα των προφίλ με τα νέα στοιχεία γίνεται μέσω της ομοιότητας συνημιτόνου.
- Το περιεχόμενο της ιστοσελίδας αναλύεται πριν από τη σύσταση και εφαρμόζεται το φιλτράρισμα περιεχομένου.
- Κάποια συστήματα εκμεταλλεύονται τις συνεργατικές τεχνικές για να βελτιώσουν τα αποτελέσματά τους.

Συστήματα recommender ηλεκτρονικού εμπορίου

- Πηγή των πληροφοριών είναι η ΒΔ των προϊόντων.
- Τα προφίλ αναπαριστώνται από την ιστορία των προϊόντων που χαρακτηρίζονται ως ενδιαφέροντα/ μη ενδιαφέροντα/ αγορασμένα.
- Τα προφίλ αυξάγονται με το χρόνο, αλλά δεν ενδιαφέρονται για τη μείωση του μεγέθους επειδή δεν θέλουν να χάσουν τις πληροφορίες.
- Δεν απαιτείται τεχνική εκμάθησης προφίλ και δεν εφαρμόζεται τεχνική προσαρμογής του για να μην ξεχαστούν τα παλαιά ενδιαφέροντα.
- Το μεγάλο μέγεθος των παραμέτρων χρήστη απαιτεί μια προηγμένη τεχνική ταιριάσματος στοιχείων-προφίλ. Μεγάλες επιχειρήσεις, όπως το Amazon και το CDNow, δεν δημοσιεύουν ποια μέθοδο χρησιμοποιούν.
- Εκμεταλλεύονται ιδιαίτερα τις συνεργατικές μεθόδους για να βελτιωσουν την ποιότητα των συστάσεων τους.

Συστήματα recommender ταινιών – μουσικής

- Εκμεταλλεύονται τις συνεργατικές μεθόδους για να βελτιώσουν τις συστάσεις τους.
- Αναπαριστούν το προφίλ χρήστη ως μήτρα προτιμήσων χρήστη-κατηγοριών δεδομένων.
- Δεν χρησιμοποιούν τεχνικές εκμάθησης προφίλ
- Ταιριάζουν τα προφίλ των χρηστών με τη μέθοδο των πλησιέστερων γειτόνων.
- Είναι σημαντικό οι χρήστες να παρέχουν ανατροφοδότηση σχετικότητας.
- Δεν ξεχνούν τα προηγούμενα ενδιαφέροντα.
- Προσαρμόζουν τις παραμέτρους χρήστη με το χρόνο, **ξεχνώντας τα παλαιά ενδιαφέροντα**.

6. ΠΑΡΑΔΕΙΓΜΑ ΑΝΑΖΗΤΗΣΗΣ ΠΛΗΡΟΦΟΡΙΩΝ

- Τα συστήματα ανάκτησης πληροφοριών (Information Retrieval-IR) επιτρέπουν στους χρήστες να εκφράσουν ερωτήσεις για να επιλέξουν
- ➡ Δεδομένα που ταιριάζουν με ένα συγκεκριμένο θέμα και εκπληρώνουν μια ιδιαίτερη ανάγκη του χρήστη για πληροφορίες στο θέμα αυτό.

Αναπαράσταση με διανύσματα

	Νουντλς Γαρίδα	Βασιλικός Σολωμός	Εξωτικός	Dolce
Karen	2.5 0	2 0	0	+
Lynn	1.1 0	1.1 1.5	0	-
Chris	1.5 0	3.5 1.5	0.5	+
Mike	1.1 1.1	21 20	25	-
Jill	1.1 22	0 0	3.5	?

Πίνακας 1 Παράδειγμα – Συνεργασία μέσω Περιεχομένου

Αναπαράσταση με διανύσματα

- Στον παραπάνω πίνακα, έχουμε τα προφίλ πέντε χρηστών που βασίζονται στο περιεχόμενο
- Επίσης την εκτίμηση των τεσσάρων πρώτων για ένα εστιατόριο, το “Dolce”.
- Υπολογίζονται οι συσχετίσεις του προφίλ της Jill με τους υπόλοιπους χρήστες.
- Χρησιμοποιώντας αυτές τις συσχετίσεις ως βάρη και την εκτίμηση των χρηστών (-1 για “-” και 1 για “+”) υπολογίζεται ο μέσος όρος τους και παρέχεται η πρόβλεψη για την Jill.

Υπολογισμός ομοιότητας (κριτήριο Pearson)

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

Where,

r = Pearson Correlation Coefficient

x_i = x variable samples

\bar{x} = mean of values in x variable

y_i = y variable sample

\bar{y} = mean of values in y variable