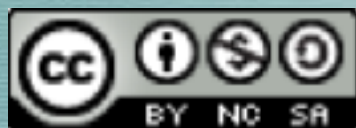




Συστήματα Παραγωγής Συστάσεων

Δημήτριος Α. Δέρβος

Τμήμα Μηχανικών Πληροφορικής & Ηλεκτρονικών Συστημάτων,
Σχολή Μηχανικών, Διεθνές Πανεπιστήμιο της Ελλάδος

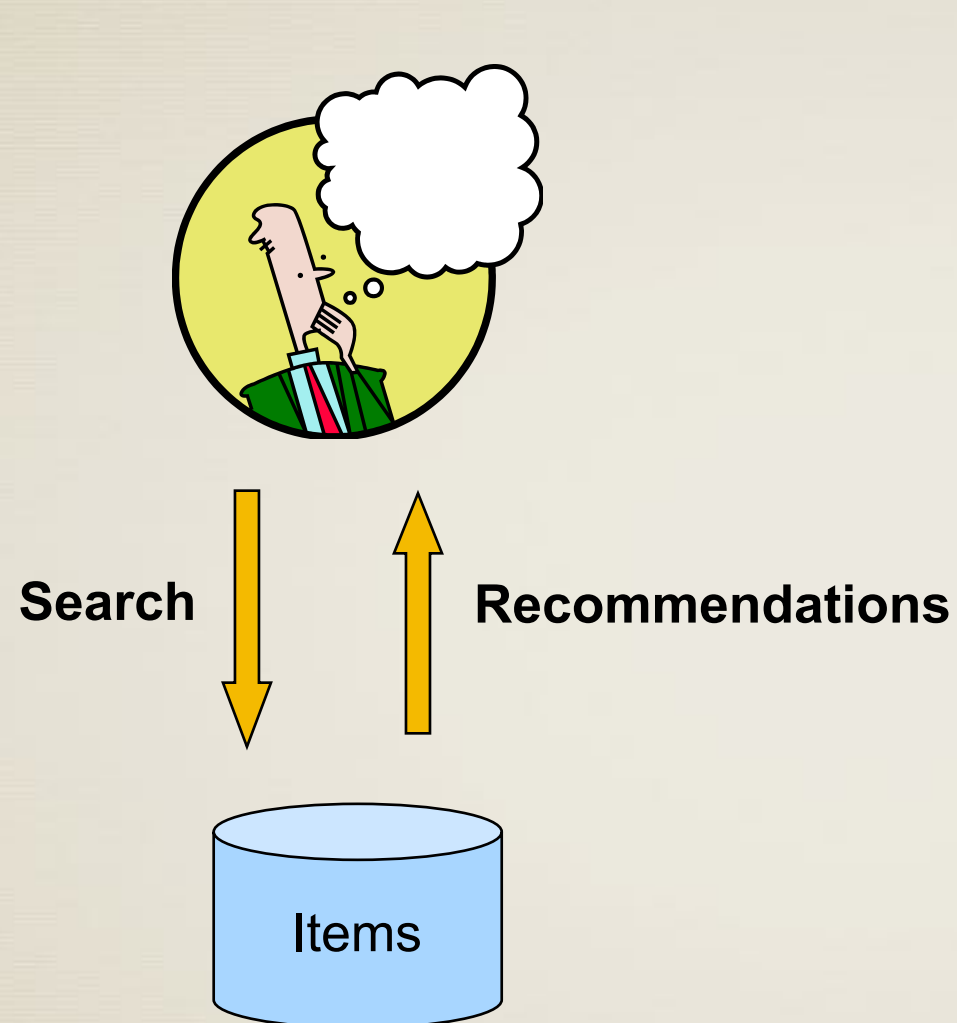


Το παρόν σετ εκπαιδευτικών διαφανειών κατοχυρώνεται με άδεια χρήσης Creative Commons (CC) Αναφοράς Δημιουργού - Μη Εμπορικής Χρήσης - Παρόμοιας Διανομής 4.0 Διεθνές (CC BY-NC-SA 4.0): <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.el>.
Η αναφορά σε τμήμα ή στο σύνολο του περιεχομένου του πρέπει να συμμορφώνεται προς τις σχετικές οδηγίες της εν λόγω άδειας χρήσης.

Αναφορές / Βιβλιογραφία

- Jiawei Han Micheline Kamber Jian Pei, Data Mining: Concepts and Techniques, 3rd Edition, 2011, Morgan Kaufmann
- Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining of Massive Datasets, 3rd Edition, Cambridge University Press, 2020, <http://www.mmds.org>
- Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, Vipin Kumar, Introduction to Data Mining, 2nd Edition, 2019, Pearson
- Mohammed J. Zaki, Wagner Meira, Jr, Data Mining and Machine Learning, Fundamental Concepts and Algorithms, 2nd edition, 2020, Cambridge University Press
- <https://www.youtube.com/watch?v=h9gpufJFF-0>

Η Ανάγκη



Από την Ανεπάρκεια στην Αφθονία

Περιορισμένη διαθεσιμότητα “χώρου” για τον τυπικό προμηθευτή
(κατάστημα, σταθμός TV, κινηματογράφος, κλπ.)

Το διαδίκτυο παρέχει άφθονο χώρο-βιτρίνα,
πρακτικά: με μηδενικό κόστος

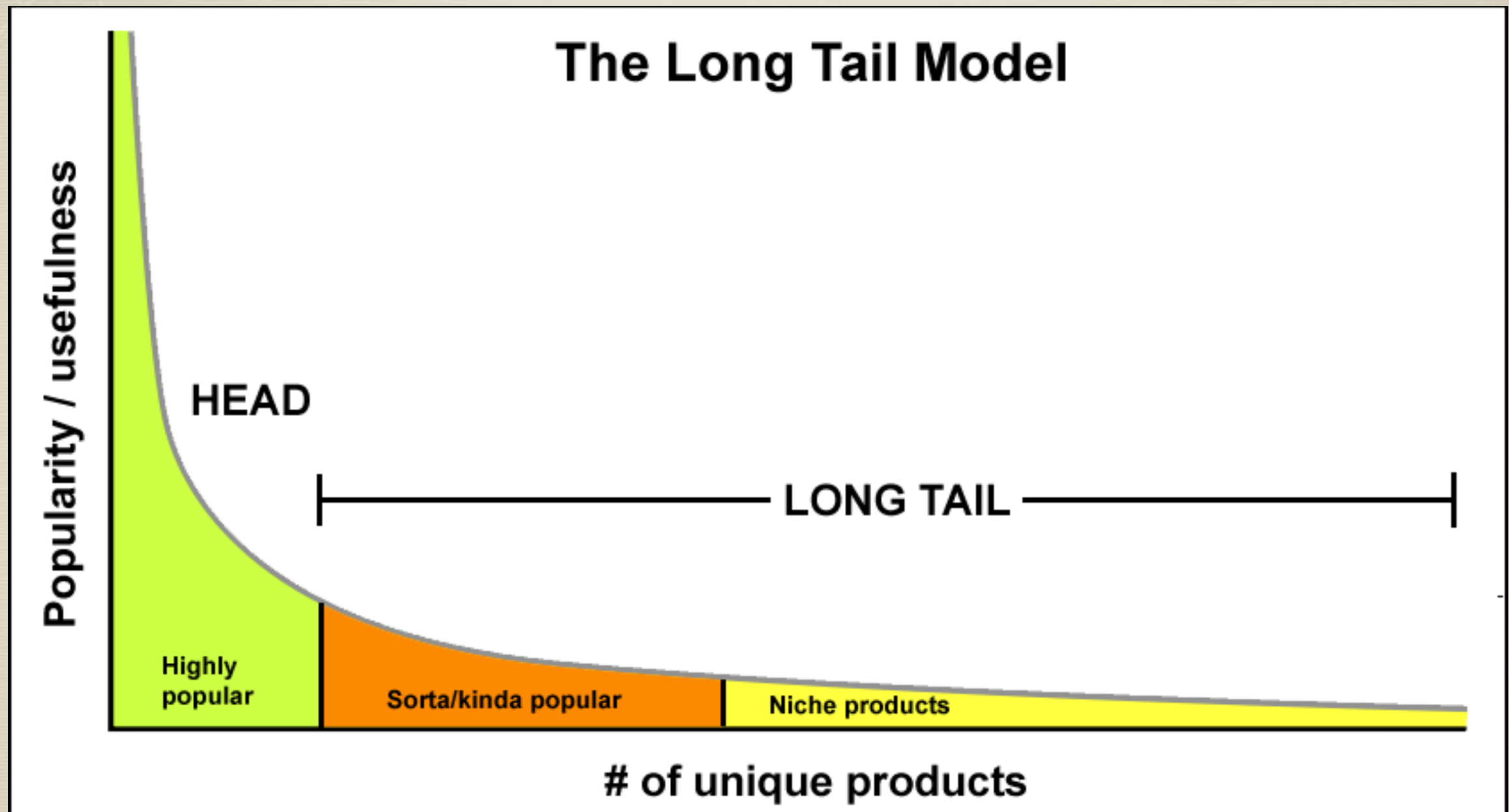
Ανάγκη “πλοήγησης” του πελάτη στον “χώρο της αφθονίας”

Η Λύση και η Οικονομία της

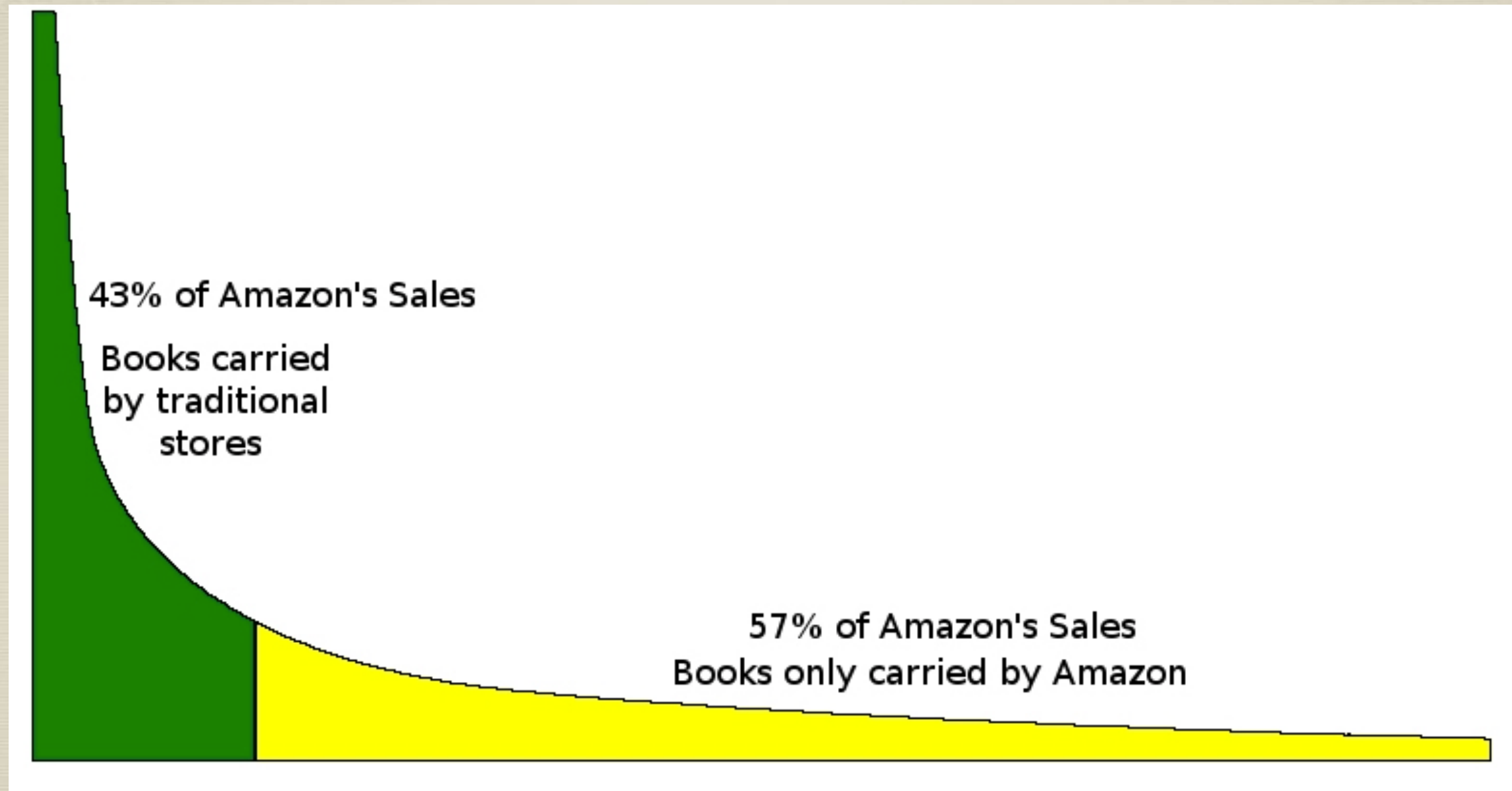
Συστήματα Παραγωγής Συστάσεων

Το βιβλίο “Into Thin Air” που έκανε το “Touching the Void” επιτυχία
(<http://www.wired.com/wired/archive/12.10/tail.html>)

Η Οικονομία της “Μεγάλης Ουράς” 1/2



Η Οικονομία της “Μεγάλης Ουράς” 2/2



Εφαρμογές

Εκεί όπου ο χρήστης δεν ξέρει “από που να αρχίσει”...

...το σύστημα του προτείνει:

- ποιά βιβλία να διαβάσει (Amazon)
- ποιές ταινίες να δει (Netflix)
- ποιούς να “κάνει φίλους” (Facebook, Linkedin, Twitter)

Απλού Τύπου Συστάσεις

Οι ειδικοί προτείνουν:

- “Αυτά που ξεχωρίζω”
- “Τα πλέον σημαντικά”

Οι καταναλωτές αναδεικνύουν:

- Top 10, 100, κλπ.
- Επισκεψιμότητα
- Downloads
- κλπ.

Στατικά, “χειρονακτικά”, χωρίς προσωποποιημένη πληροφόρηση

Εξελιγμένες Συστάσεις

Λαμβάνονται υπόψιν τα ιδιαίτερα χαρακτηριστικά και οι προτιμήσεις του χρήστη:

Βιβλία ανάλογα με αυτά που συνήθως προτιμά

Ιστοσελίδες με περιεχόμενο που τον ενδιαφέρει

Ταινίες του είδους που του αρέσουν

Ταξιδιωτικοί προορισμοί που έχουν ενθουσιάσει τους φίλους του

Τραπεζικά προϊόντα που αγοράζουν πελάτες με το δικό του προφίλ

Μοντελοποίηση των Προτιμήσεων

X = σύνολο πελατών

S = σύνολο προϊόντων

Συνάρτηση οφέλους (utility function) $u: X \times S \rightarrow R$

R = σύνολο αξιολογήσεων:

0-5 αστέρια, [Αρέσει, Δεν Αρέσει], κλπ.

Πίνακας Οφέλους (Utility Matrix) 1/3

	TRUE BELIEVER	THE DA VINCI CODE	THE WORLD IS FLAT	MY LIFE SO FAR	THE TAKING	THE KITE RUNNER	RUNNY BABBIT	HARRY POTTER
U1	1	5		3			3	5
U2	5	4			3	2	1	
U3	3		1	2	2			5
U4		3			4	1		3
U5	2	4	3			2	2	
U6	5			3	1		3	1
U7	1	4	5	5	2			4
U8	2	1			4	5	1	
U9			3	2	2			5
U10	3	5	1				4	4
U11			2	1		2		3
U12	4	4		2		1	1	4
U13			2		4		4	5
U14		5	3	3	2		1	1
U15		2			3	3		2
U16		3	2	1	1		4	4
U17	1	5	1	2		4		4
U18	5		4		3	3	4	5
U19		4		2		5	1	5
U20	2	5	1	1	5	3		4

Πίνακας Οφέλους (Utility Matrix) 2/3

Συνήθως: αραιός

Διαχείριση των κενών κελιών

Παρέκταση (extrapolation, “πρόβλεψη”) τιμών για τα κενά κελιά:
υψηλή τιμή αξιολόγησης προϊόντος από πελάτη; -> Σύσταση!

Πίνακας Οφέλους (Utility Matrix) 3/3

Ζητήματα προς αντιμετώπιση:

1. Συλλογή δεδομένων αξιολόγησης
2. Παρέκταση (extrapolation) προς τις τιμές των κενών κελιών
3. Για το (2): ενδιαφέρουν οι υψηλές τιμές αξιολόγησης
4. Αξιολόγηση της ποιότητας της παρέκτασης

Συλλογή Δεδομένων Αξιολόγησης

Άμεσος (explicit) τρόπος:

- ρωτώντας τους χρήστες/πελάτες
- προτερήματα: απλή διαδικασία, άμεση επικοινωνία με τους πελάτες - αξιόπιστα δεδομένα αξιολόγησης
- μειονέκτημα: αδυναμία κλιμάκωσης: μικρός ο αριθμός αυτών που ανταποκρίνονται

Έμμεσος (implicit) τρόπος:

- επεξεργασία και ερμηνεία ενεργειών των πελατών (αγορά = υψηλή προτίμηση)
- εφικτή η κλιμάκωση
- πως οι χαμηλές προτιμήσεις;

Στην πράξη: συνδυασμός των παραπάνω δύο

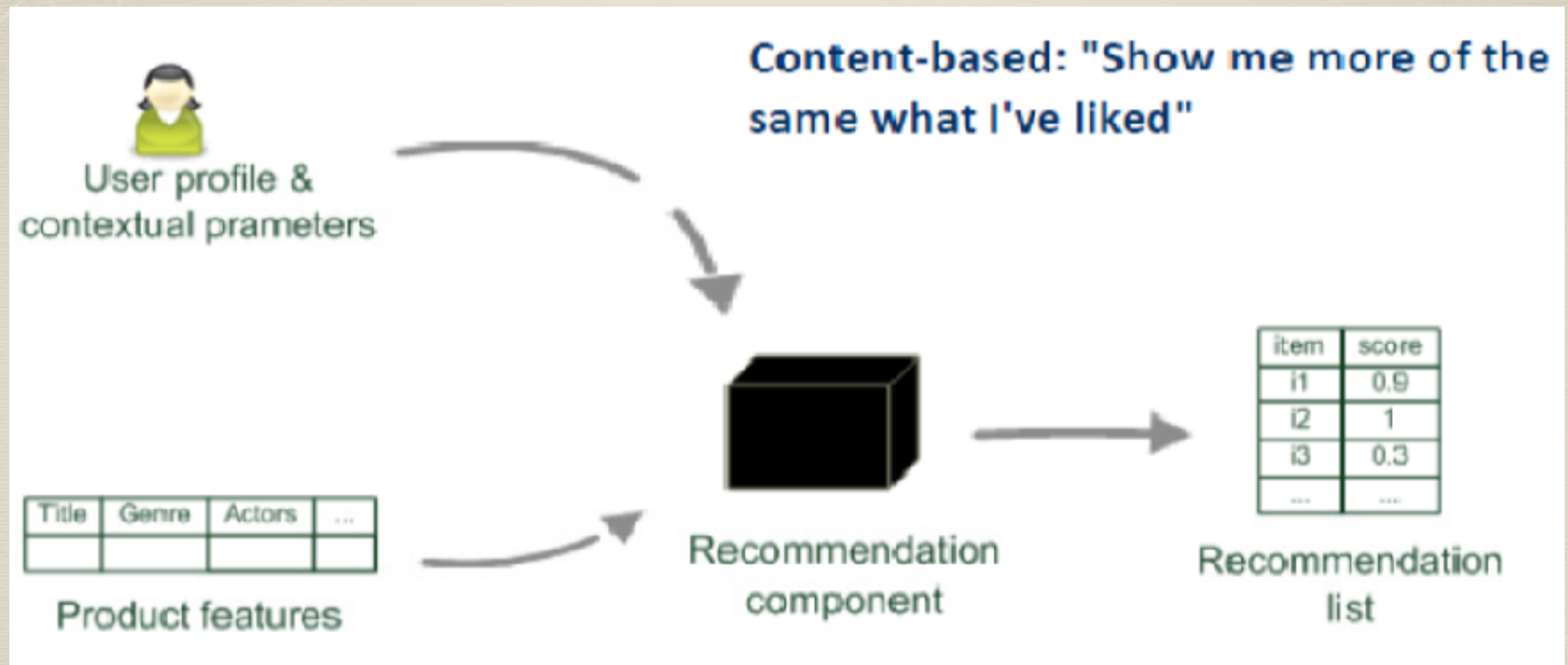
Μεθοδολογίες Παρέκτασης

- πίνακας οφέλους: αραιός (πολλοί δεν αξιολογούν πολλά)
- “ψυχρή εκκίνηση”: πελάτης χωρίς προϋστορία προτιμήσεων, προϊόν που κανείς δεν το έχει αξιολογήσει, ακόμη

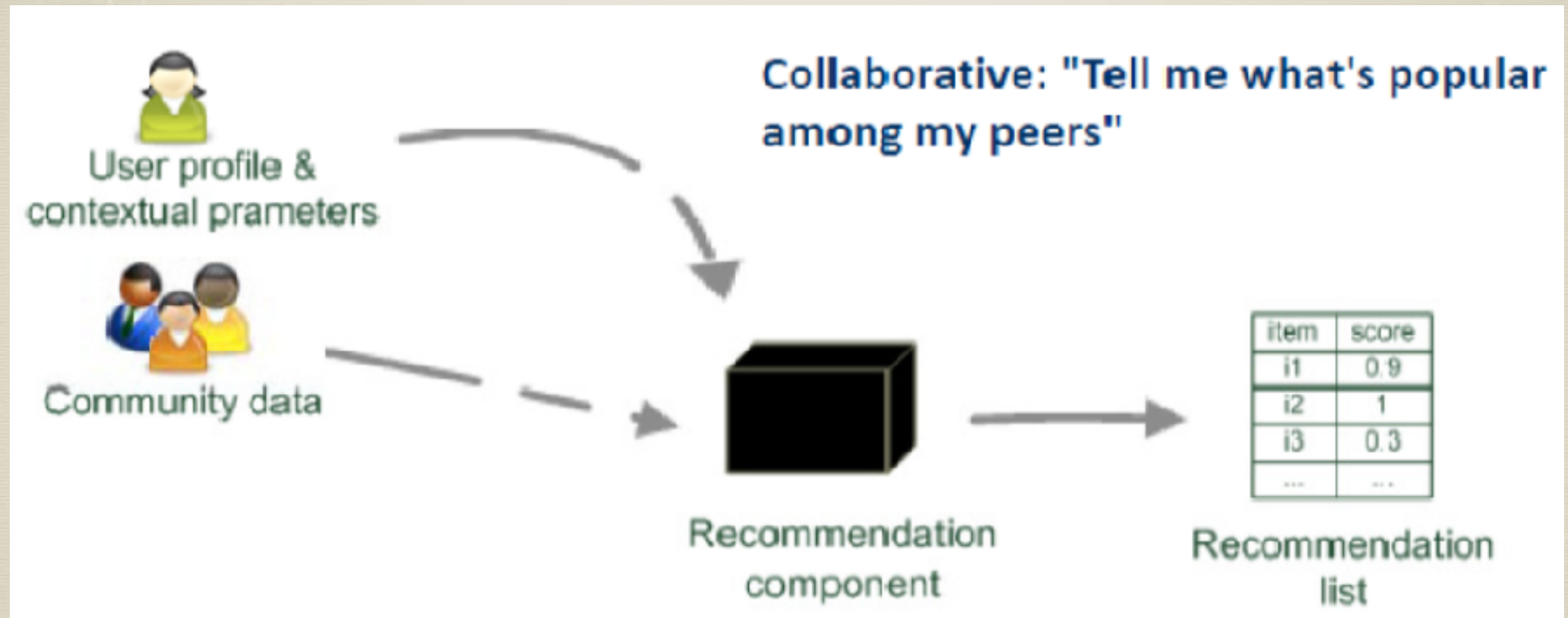
Τρεις (βασικοί) τρόποι προσέγγισης:

1. Βάσει περιεχομένου (content-based)
2. Βάσει συνεργατικών φίλτρων (collaborative filtering)
3. Υβριδικός (συνδυασμός των παραπάνω δύο, και όχι μόνον)

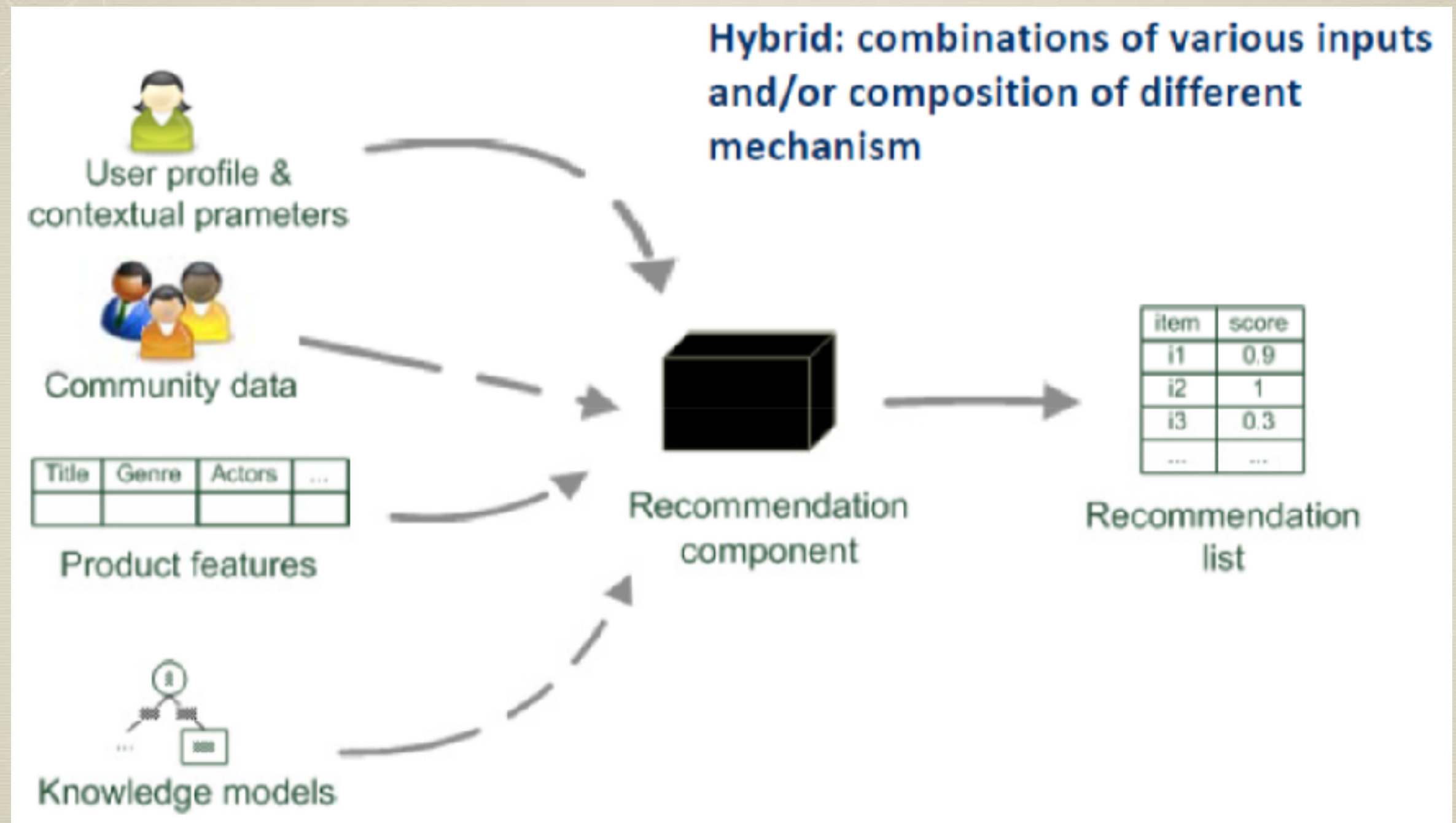
Βάσει του Περιεχομένου (Content-Based)



Βάσει Συνεργατικών Φίτρων (Collaborative Filtering)



Υβριδική Προσέγγιση



Τέλος

Ερωτήσεις;