

# Metódy získavania používateľských preferencií

Veronika Vaneková

## Modely preferencií

### Vektor ohodnotení objektov

**Reprezentácia:** Vektor  $(p_1, \dots, p_n)$ , kde  $p_i$  je ohodnotenie i-teho výrobku. Tento vektor je typicky riedky, teda väčšina hodnôt je neznámych.

**Využitie:** Ľubovoľná doména. Systém odhadne ohodnotenie objektu, ktorý používateľ ešte nevidel a odporúča objekty s najvyšším odhadovaným ohodnotením [1].

**Získavanie:**

- priamo od používateľa
- kolaboratívne filtrovanie podľa používateľov: systém nájde skupinu podobných používateľov podľa ich vektorov preferencií. Potom pre každý výrobok spočíta priemerné ohodnotenie v rámci skupiny a vyberie objekty s najvyšším ohodnotením.
- kolaboratívne filtrovanie podľa objektov: systém určí skupinu podobných objektov a na základe priemerných ohodnotení v tejto skupine ich odporúča používateľovi.

**Výhody:** Jednoduchý model, dostatočná presnosť v praxi.

**Nevýhody:** Nevieme určiť preferencie pre nového používateľa, ktorý ešte neohodnotil nijaké objekty [5].

### Content-based filtering

**Reprezentácia:** Množina kľúčových slov.

**Využitie:** Doména filmov, kníh, článkov, ... (objekty sú v textovom formáte alebo sú textovo anotované). Systém odporúča používateľovi tie objekty, ktoré sa s profilom najviac zhodujú (tf-idf model, clustering).

**Získavanie:**

- priamo od používateľa
- z ohodnotených objektov (data mining, štatistické učenie) [2]

**Výhody:** Minimalistický model.

**Nevýhody:** Malá presnosť, problém určenia preferencií pre nového používateľa [5].

### Multikriteriálne rozhodovanie

**Reprezentácia:** Systém eviduje rôzne vlastnosti (atribúty) objektov. Používateľ môže vybrať množiny alebo intervaly hodnôt vlastností, ktoré ho zaujímajú.

**Využitie:** Elektronický obchod, hlavne doména notebookov.

**Získavanie:**

- priamo od používateľa cez vstupný formulár
- indukčnými metódami z ohodnotených objektov

**Výhody:** Detailná reprezentácia preferencií.

**Nevýhody:** Objektov môže byť príliš veľa, nie je možné ich usporiadať podľa relevancie. Pri konjunktívnom dopyte systém nezíská tie objekty, ktorým chýba jedna požadovaná vlastnosť, ale pritom spĺňajú ostatné vlastnosti [3].

### Riešenie: Fuzzy model

**Reprezentácia:** Použijeme fuzzy množiny hodnôt, ktoré pripúšťajú čiastočné členstvo prvkov. Hodnota charakteristickej funkcie je z intervalu  $[0,1]$  a vyjadruje preferencie používateľa k nejakej vlastnosti objektu.

Celková relevancia objektu podľa viacerých vlastností sa vypočíta cez agregáčnu funkciu. Príkladom je vážený priemer alebo monotónna klasifikácia, ktorá sa dá získať z dát a zapísať vo forme pravidiel [4].

**Využitie:** Ľubovoľná doména, napríklad pracovné ponuky.

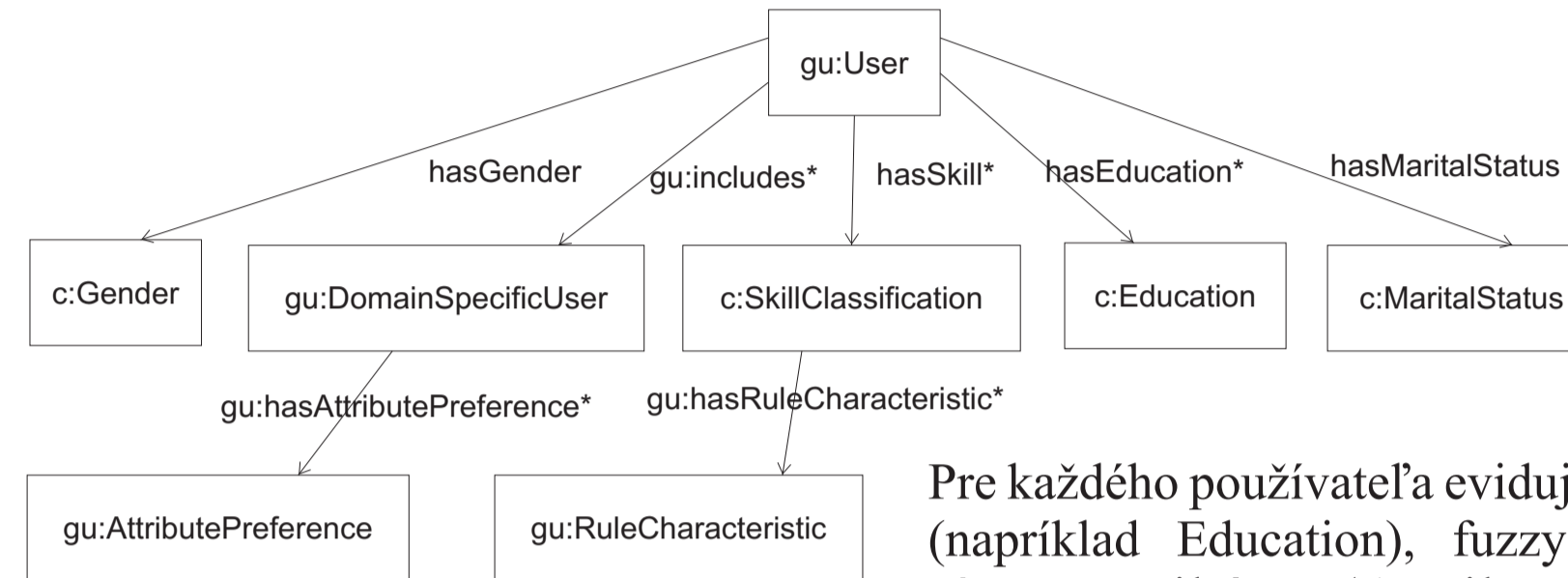
**Získavanie:**

- priamo od používateľa cez GUI
- indukčnými metódami z ohodnotených objektov

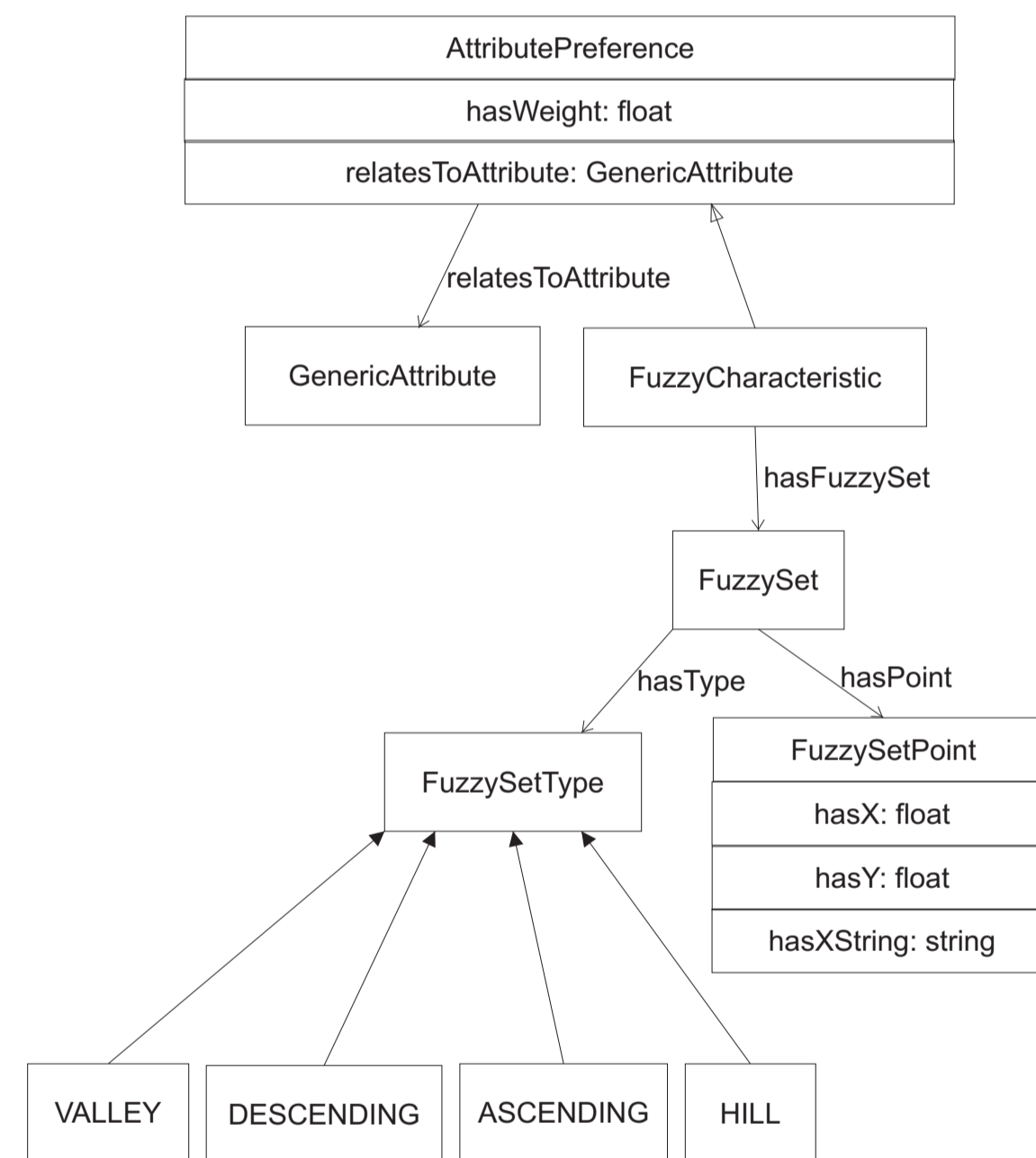
**Výhody:** Detailná reprezentácia preferencií.

**Nevýhody:** Menej intuitívne.

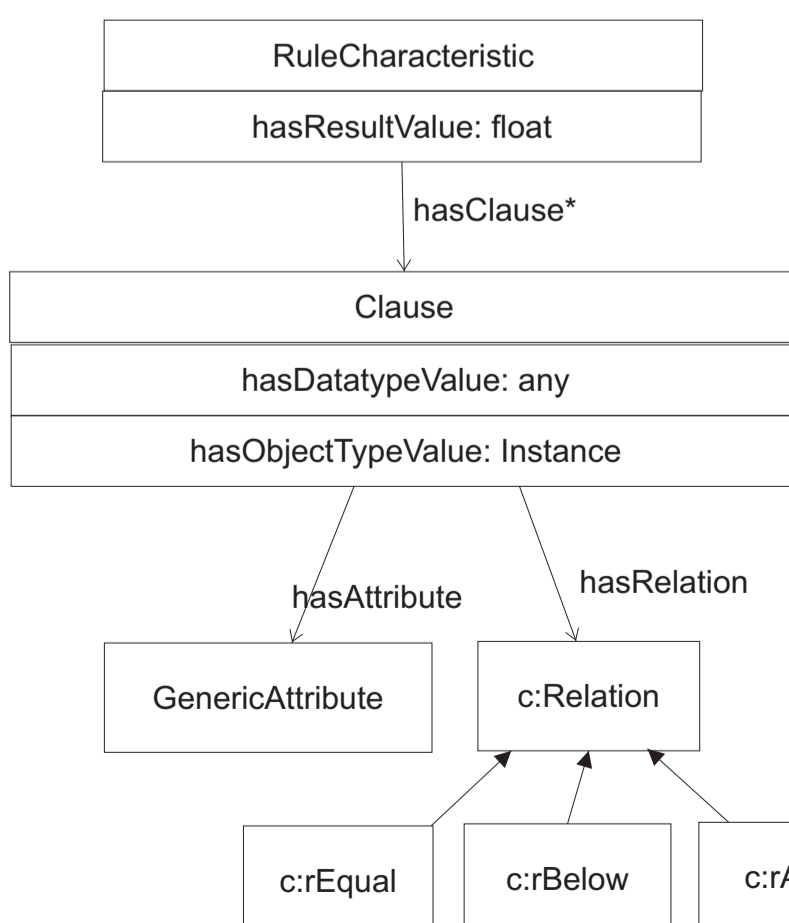
## Ontologický model



Pre každého používateľa evidujeme osobné údaje (napríklad Education), fuzzy preferencie pre rôzne atribúty (AttributePreference) a klasifikačné pravidlá (RuleCharacteristic).



Preferencie sa vzťahujú k jednotlivým atribútom. Každá inštancia obsahuje fuzzy množinu, ktorá je jednoznačne určená konečnou množinou bodov. Medzi týmito bodmi má charakteristická funkcia lineárny priebeh. Podľa tvaru tejto funkcie patrí fuzzy množina do jedného zo štyroch základných typov (ASCENDING, DESCENDING, HILL, VALLEY).

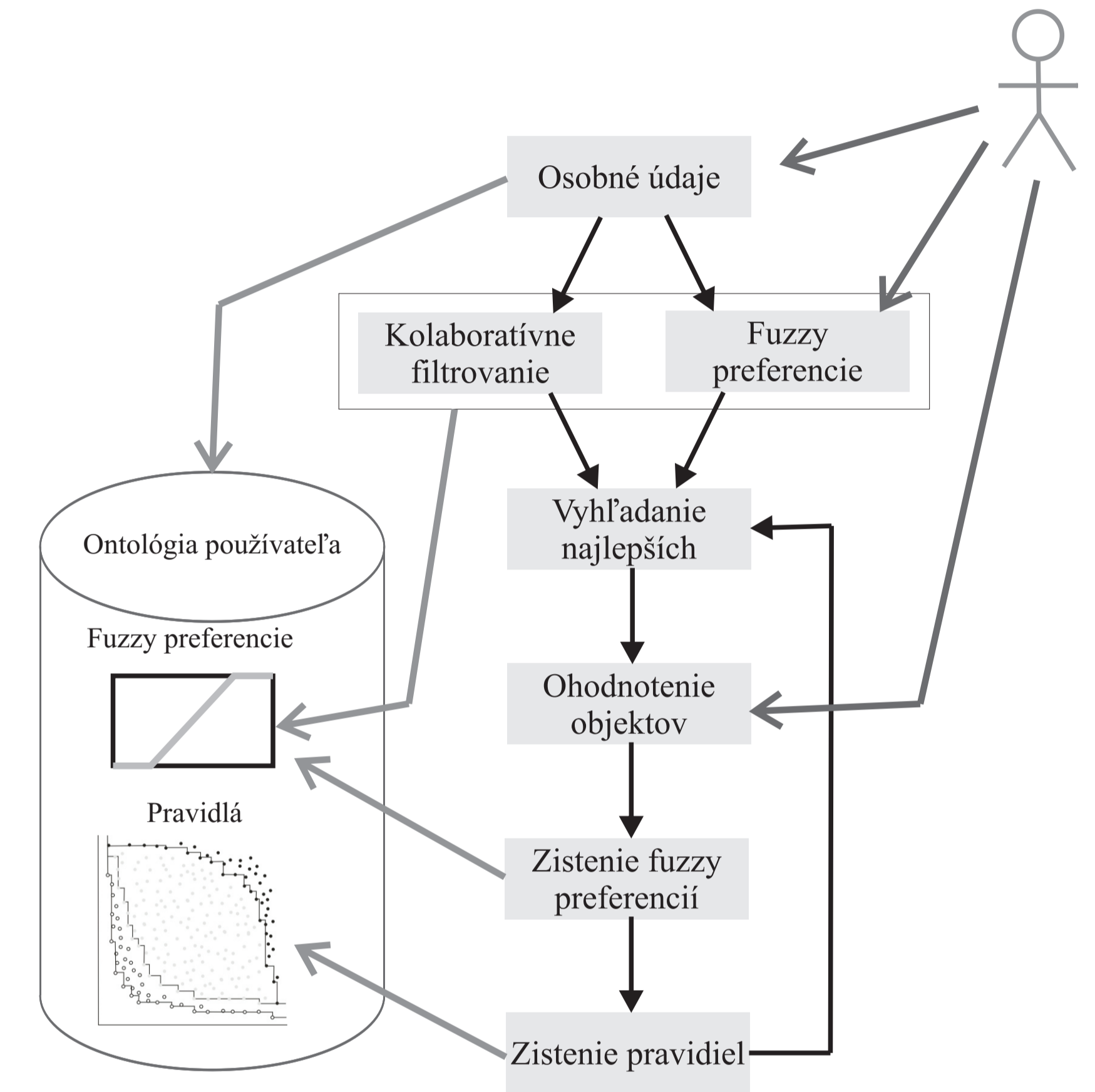


Klasifikačné pravidlá skombinujú hodnoty fuzzy preferencií do jedného výsledku. Každé pravidlo pozostáva z niekoľkých podmienkových klauzúl (Clause) a výslednej hodnoty. Ak sú klauzuly splnené, tak výsledná hodnota objektu bude väčšia alebo rovná ako výsledná hodnota pravidla.

Príklady pravidiel:

- $\text{dobry\_notebook}(x) \geq 0.8$  IF  $\text{dobra\_cena}(x) \geq 0.8$  AND  $\text{dobra\_ram}(x) \geq 0.5$
- $\text{dobry\_notebook}(x) \geq 0.4$  IF  $\text{dobry\_disk}(x) \geq 0.5$  AND  $\text{dobra\_uhlopriecka}(x) \geq 0.8$
- $\text{dobry\_notebook}(x) \geq 0.4$  IF  $\text{dobra\_cena}(x) \geq 0.6$

## Správa preferencií



Podľa toho, koľko informácií o používateľovi máme a aké nové informácie dostaneme, sa aktualizujú fuzzy preferencie a agregáčna funkcia v ontológii. Typ vstupu je špecifikovaný v hlavičke každého stĺpca.

	Osobné údaje	Explicitné fuzzy množiny	Implicitné obmedzenia atribútov	Ohodnotenie vzorky objektov
Fuzzy množiny	Kolaboratívne filtrovanie	Prepíše	Prispôbi podľa obmedzení	Indukciou odvodí nové fuzzy množiny.
Agregáčna funkcia	Nastaví rovnaké váhy	Nastaví rovnaké váhy	-	Indukciou odvodí pravidlá.

- [1] Candillier, L., Meyer, F., Boullé, M. Comparing State-of-the-Art Collaborative Filtering Systems. Machine Learning and Data Mining in Pattern Recognition. Springer Berlin / Heidelberg 2007. ISBN 978-3-540-73498-7.
- [2] Grossman, D. A., Frieder, O.: Information Retrieval: Algorithms and Heuristics, Kluwer Academic Publishers, 2000. ISBN 1402030037.
- [3] Gurský, P.: Towards better semantics in the multifeature querying. Proceedings of DATESO 2006.
- [4] Horváth, T., Vojtáš, P.: Ordinal Classification with Monotonicity Constraints. Proceedings of ICDM 2006. Springer, 2006. ISBN 0-7695-2702-7.
- [5] Schein, A. I., Popescul, A., Ungar, L. H., Pennock, D. M. Methods and metrics for cold-start recommendations. Proceedings of the 25th Annual international ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York 2002. ISBN 1-58113-561-0.
- [6] Tvarožek M., Bieliková M.: Adaptive faceted browser for navigation in open information spaces. Proceedings of the 16th international conference on World Wide Web. ACM Press, New York, USA. ISBN 978-1-59593-654-7.