

# Doporučování multimediálního obsahu s využitím senzoru Microsoft Kinect

Jaroslav Kuchar<sup>1,2</sup> and Tomáš Kliegr<sup>2</sup>

<sup>1</sup> Web Engineering Group, Faculty of Information Technology,  
Czech Technical University in Prague  
`firstname.lastname@fit.cvut.cz`

<sup>2</sup> Dep. of Information and Knowledge Engineering,  
Faculty of Informatics and Statistics, University of Economics Prague  
`firstname.lastname@vse.cz`

**Abstrakt** Tento příspěvek představuje online recommender InBeat.eu. Systém umožňuje sběr explicitních a implicitních zpětných vazeb od uživatelů, které jsou použity jako ukazatele zájmu. Demonstrace systému je zaměřena na interakci uživatelů s multimediálním obsahem, konkrétně se jedná o videa a scénář televizních zpráv. Video jsou automaticky sémanticky anotována s pomocí nástroje pro hledání pojmenovaných entit. Důležitou součástí systému je propojení se senzorem Microsoft Kinect, který umožňuje analyzovat natočení hlavy za účelem reálného vyhodnocení sledování daného videa. Ze získaných dat jsou odvozena asociační pravidla představující preference daného uživatele. Tyto pravidla jsou následně použita pro doporučení.

**Keywords:** doporučení obsahu, multimediální obsah, sémantika, učení preferencí, asociační pravidla

## 1 Úvod

Existují dva druhy zpětné vazby, které lze využít jako vstup při učení preferencí a následném doporučení obsahu: explicitní a implicitní zpětná vazba. Explicitní vyžaduje určitou akci uživatele aby vyjádřil úroveň zájmu, např. uložení do oblíbených položek či ohodnocení videa. Jako implicitní zpětnou vazbu lze využít např. samotné spuštění videa nebo čas strávený při sledování videa. Důležitou složkou implicitní zpětné vazby může být fyzické chování uživatele: např. do jaké míry opravdu dané video uživatel sleduje během jeho přehrávání. Pro získání tohoto typu zpětné vazby jsou využívány senzory sledování pohybu očí [8], které ale nejsou běžně dostupné. Alternativou je sledování pohybu celé hlavy. Ze směru natočení hlavy lze odvodit, zda uživatel sleduje obrazovku či nikoliv.

InBeat.eu je obecnou webovou službou pro sběr interakcí a učení preferencí, která je v současné době rozvíjena pro doporučení obsahu uživatelům chytrých televizí [4]. Tento příspěvek popisuje novou verzi InBeat, která využívá běžně dostupného senzoru Microsoft Kinect pro sledování pohybu celé hlavy. Ze směru natočení hlavy se odvozuje, zda uživatel sleduje obrazovku či nikoliv. Naším předpokladem je, že natočení hlavy je dostatečnou informací pro

stanovení zájmu. Jelikož byl Microsoft Kinect navržen pro sledování pohybu těla v prostoru, není schopen spolehlivě poskytovat žádnou detailnější informaci, např. mimiku obličeje.

## 2 Scénář demonstrace

V této části je popsán základní scénář pro demonstraci. Zahrnuje různé druhy akcí osoby a s tím související zpětná vazba, která je následně použita pro učení preferencí.

Uživatel dostane možnost pustit přehrávač simulující rozhraní televize. V rámci demonstrace je předpřipravena sada krátkých videí představujících televizní zprávy. Televizní zprávy obvykle obsahují vstupy z různých oblastí jako jsou politika, sport, počasí či různé zajímavosti. Jako ukázkovou osobu mějme uživatele zájímajícího se o sport, konkrétně fotbalového fanouška.

*Televizní zprávy začínají sadou vstupů o politice. Uživatel je pouze sleduje, některé části přeskakuje. Jakmile se uživatel dostane ke sportovním událostem, začne sledovat pozorněji. V okamžiku zprávy o oblíbeném fotbalistovi si ukládá tuto část videa do oblíbených položek aby se mohl k tomuto videu později vrátit. Uživatel má také možnost si video pozastavit a přečíst si doplňující informace pomocí připojeného odkazu. Dalším vstupem jsou informace o možnostech návštěvy Berlína a jeho památek. Tato část videa uživatele nezajímá a proto začne hrát hru na tabletu.*

Všechny akce i zpětné vazby jsou zaznamenány pomocí systému a transformovány na ukazatele zájmu s různou vahou. Po skončení televizních zpráv má systém dostatečné množství dat k naučení pravidel vyjadřujících jeho preference. Příkladem takových pravidel jsou:

*“SoccerPlayer(yes)  $\rightarrow$  interest(positive)”*,  
*“Berlin(yes)  $\wedge$  Politics(yes)  $\rightarrow$  interest(neutral)”*.

Tato pravidla jsou součástí profilu daného uživatele a jsou použita k doporučení dalších videí či souvisejícího obsahu.

## 3 Popis dílčích kroků

V této části jsou popsány dílčí kroky, které se dějí na pozadí při ukázce systému.

**Analýza titulků.** Důležitým vstupem pro celou službu je sémantický popis každého přehrávaného videa. InBeat na pozadí zpracovává dostupná metadata v podobě titulků nebo automatické transkripce a posílá je k analýze do systému pro rozpoznávání pojmenovaných entit `entityclassifier.eu` [1]. Video je rozděleno na menší jednotky odpovídající délce jednoho titulku. Výstupem je seznam detekovaných entit pro každý jednotlivý titulek. Každá entita sebou nese informaci o svém typu, který je popsán pomocí typů DBpedia. Jsou použity technologie sémantického webu a Linked Data pro popis každého titulku.

**Sběr interakcí.** Pro účely demonstrace je použit upravený přehrávač YouTube s vlastním ovládáním. Všechny akce určené pro ovládání přehrávače i výsledky

sledování chování pomocí Microsoft Kinect [6] jsou posílány na server InBeat k následnému zpracování.

**Agregace interakcí.** Jelikož během doby zobrazení jednoho titulku může dojít k několika interakcím, jsou interakce sloučeny do jednoho výsledného ukazatele zájmu. Pro agregaci jsou použita jednoduchá experimentálně definovaná pravidla. Např. pokud se uživatel dívá na obrazovku, zvyš hodnotu zájmu o 0,3 nebo v případě pokud uživatel přeskočil část videa, sniž hodnotu o 0,2. Tato pravidla byla stanovena experimentálně a správné nastavení je předmětem dalších experimentů a evaluace. Výsledná hodnota je normalizována na interval  $(-1, 1)$  a převedena na jednu z těchto tříd: negativní  $(-1, 0)$ , neutrální  $(0, 0)$ , pozitivní  $(0, 1)$ .

Identifikace			Popis		
userId	sessionId	pseudo shotId	Neymar	SoccerPlayer	... Interest
user1	1124541	125	yes	yes	... positive

**Tabulka 1.** Ukázka formátu pro učení asociačních pravidel.

**Učení preferencí.** Pro učení asociačních pravidel a preferencí uživatele je použit systém EasyMiner ([easyminer.eu](http://easyminer.eu)). Vstupem pro učení asociačních pravidel je tabulkový formát složený především z popisu jednotlivých titulků a výsledné kategorie vyjadřující úroveň zájmu. Příklad je uveden v Tabulce 1. První část definuje identifikátory uživatele, identifikátor sezení a identifikátor pseudo záběru, což je fragment videa, která odpovídá době zobrazení titulku. Druhou částí je popis fragmentu videa pomocí pojmenovaných entit a jejich typů. Poslední část obsahuje výslednou hodnotu míry zájmu, která vznikla agregací interakcí zachycených během pseudo záběru. Pro samotné učení preferencí v podobě asociačních pravidel je použita druhá část obsahující popis pro levou stranu pravidel a míra zájmu pro pravou stranu pravidel.

**Doporučování.** Vstupem pro doporučování jsou pravidla představující preference uživatelů a seznam kandidátů pro doporučení. Seznam kandidátů může být stanoven např. pomocí jednoduchého vyhledávání pomocí klíčových slov. Pro každého kandidáta jsou nalezena asociační pravidla. Nejsilnější z nalezených pravidel je použito k přiřazení úrovně zájmu. Síla pravidla je určena pomocí upraveného klasifikačního algoritmu CBA (Classification Based on Associations)[3]. Následně jsou kandidáti seřazeni podle přiřazené úrovně zájmu a tento seznam představuje doporučení pro uživatele.

## 4 Závěr

Hlavním přínosem prezentovaného systému je:

- InBeat.eu je, pokud víme, prvním on-line systémem pro doporučování videí používajícím běžně dostupný senzor Microsoft Kinect pro stanovení úrovně zájmu uživatele.
- Využití systému pro rozpoznávání pojmenovaných entit a provázání s Linked Data za účelem sestavení popisu videa. Analýza je provedena jak pro videa,

kteřá uživatel sleduje tak i pro videa určená k doporučení. Tím je zajištěna jejich tématická souvislost.

Díličí části prezentovaného systému byly experimentálně ověřeny, včetně využití Microsoft Kinect pro účely detekce pozornosti uživatele při sledování videa [7]. Do budoucnosti je v plánu podrobná evaluace celého systému [5]. Anglická verze článku byla přijata na konferenci PAIS 2014, ECAI [2].

### Poděkování.

Náš dík mají kolegové Julien Leroy a Matei Mancias z University of Mons, kteří nám poskytli software pro sledování uživatelů pomocí Microsoft Kinect. Tato práce byla podpořena grantem Studentské grantové soutěže ČVUT č. SGS14/104/OHK3/1T/18, grantem IGA 26/2011 Vysoké školy ekonomické, Praha a projektem FP7-287911 LinkedTV.

### English Summary

This paper demonstrates Interest Beat (InBeat.eu) as a recommender system for online videos, which determines user interest in the content based on gaze tracking with Microsoft Kinect in addition to explicit user feedback. Content of the videos is represented using a semantic wikifier. User profile is constructed from preference rules, which are discovered with an association rule learner.

### Reference

1. M. Dojchinovski and T. Kliegr. Entityclassifier.eu: real-time classification of entities in text with Wikipedia. In *ECML'13*, pp. 654–658. Springer.
2. T. Kliegr and J. Kuchař. Orwellian eye: Video recommendation with microsoft kinect. In *ECAI*. 2014.
3. T. Kliegr, J. Kuchař, D. Sottara, and S. Vojíř. Learning business rules with association rule classifiers. In A. Bikakis, P. Fodor, and D. Roman, (eds.) *Rules on the Web. From Theory to Applications*, vol. 8620 of *Lecture Notes in Computer Science*, pp. 236–250. Springer International Publishing, 2014.
4. J. Kuchař and T. Kliegr. Gain: Web service for user tracking and preference learning - a smart tv use case. In *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, pp. 467–468. ACM, New York, NY, USA, 2013.
5. J. Kuchař and T. Kliegr. Bag-of-entities text representation for client-side recommender systems. In *First Workshop on Recommender Systems for Television and online Video (RecSysTV)*, *ACM RecSys*. 2014.
6. J. Leroy, F. Rocca, M. Mancias, and B. Gosselin. 3d head pose estimation for TV setups. In *Intelligent Technologies for Interactive Entertainment*, vol. 124, pp. 55–64. Springer, 2013.
7. J. Leroy, *et al.* KINterestTV - towards non-invasive measure of user interest while watching tv. In Y. Rybarczyk, T. Cardoso, J. Rosas, and L. Camarinha-Matos, (eds.) *Innovative and Creative Developments in Multimodal Interaction Systems*, vol. 425 of *IFIP Advances in Information and Communication Technology*, pp. 179–199. Springer Berlin Heidelberg, 2014.
8. S. Xu, H. Jiang, and F. C. Lau. Personalized online document, image and video recommendation via commodity eye-tracking. In *RecSys '08*, pp. 83–90. ACM, New York, NY, USA, 2008.