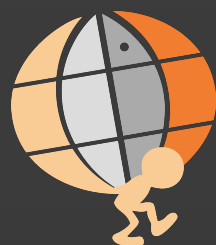


Prepájanie multimediálnych metadát

DP III.

Bc. Peter Gašpar
gasparpeter92@gmail.com

Vedúci: Ing. Jakub Šimko, PhD.



PeWe@FIIT
personalized web group

31.3.2016

DOMÉNA MULTIMÉDIÍ

- problémová doména:
TV program (a jeho entity)
- súčasný stav:
dominujú **statické informácie**
- zdroj dynamických informácií:
používatelia a ich aktivita na **sociálnych sieťach**



SOCIÁLNE SIETE

- z pohľadu TV vysielania:
- verejné príspevky používateľov,
- **interakcia medzi televíziou a divákom,**
- diskusia medzi divákmi,
- na Slovensku dominuje **Facebook**, vo svete Twitter.



tvmarkizaofficial

SLEDOVAŤ

1,132 Páči sa mi to

2 d

tvmarkizaofficial ste pripraveni na dnesne #semifinale dlevecat? #superstar2015 kto sa dostane do #finale sledujte #dnes #vecer 20:25 #tvmarkiza

63%
opakované
reakcie ¹⁾

1) percento príspevkov, v ktorých aspoň 1 používateľ pridal aspoň 2 komentáre,

Analyzovaná stránka:
facebook.com/csmatalent, máj
- október 2015

HYPOTÉZY A CIELE

TV PROGRAM

17:00 Televízna relácia 1

18:00 Televízna relácia 2

19:00 Televízna relácia 3

20:00 Televízna relácia 4

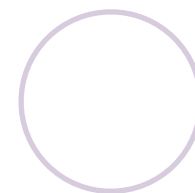
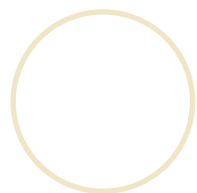
21:00 Televízna relácia 5

22:00 Televízna relácia 6

?

SOCIÁLNA SIEŤ

Budete dnes večer pozerať
Televíznu reláciu 3?



HYPOTÉZY

TV PROGRAM

17:00 Televízna relácia 1

18:00 Televízna relácia 2

19:00 Televízna relácia 3

20:00 Televízna relácia 4

21:00 Televízna relácia 5

22:00 Televízna relácia 6

?

SOCIÁLNA SIEŤ

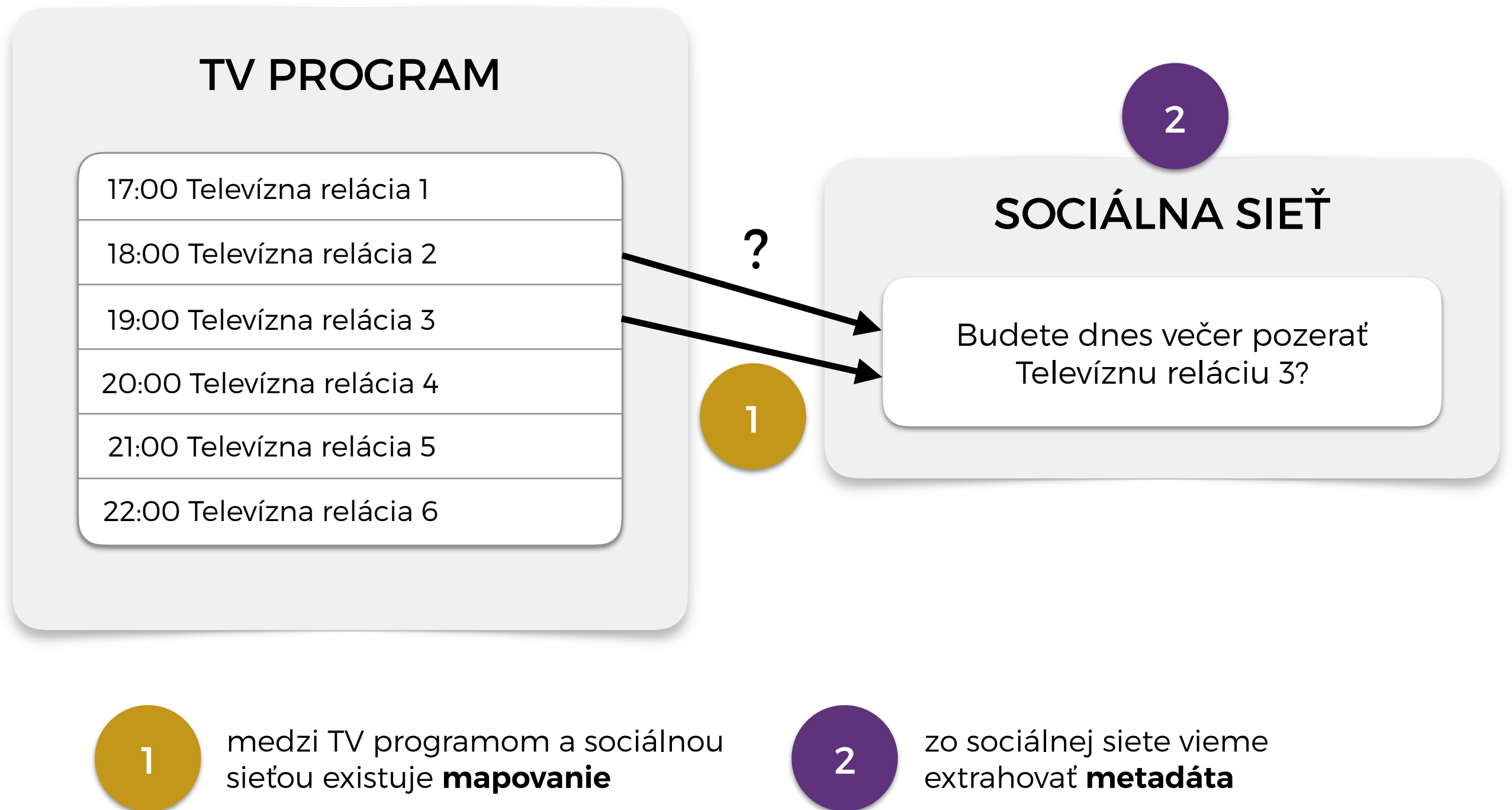
Budete dnes večer pozerať
Televíznu reláciu 3?

1

1

medzi TV programom a sociálnou
sieťou existuje **mapovanie**

HYPOTÉZY



PREHL'AD SÚČASNÉHO STAVU

- základ: **báza znalostí**
 - Wikipedia, DBPedia, IMDB, CSFD
 - **TV Program**
- mapovanie:
 - základná črta príspevku: **text príspevku, hashtagy**
 - n-gramy, TF-IDF, rôzne metódy strojového učenie

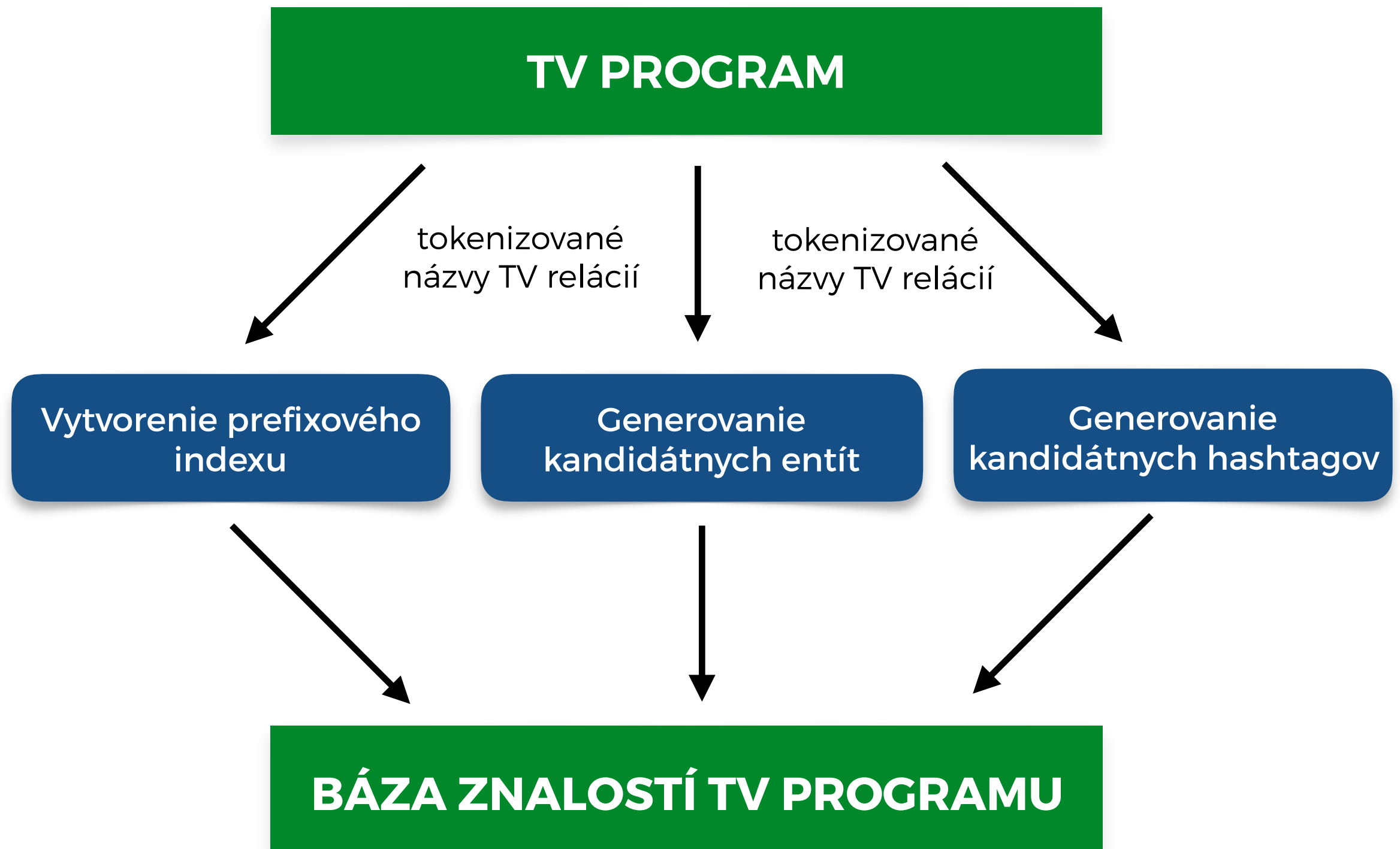


OBSAH ZO SOCIÁLNEJ SIETE



NAŠA HYPOTÉZA:
Podobné črty vieme získať z TV programu.

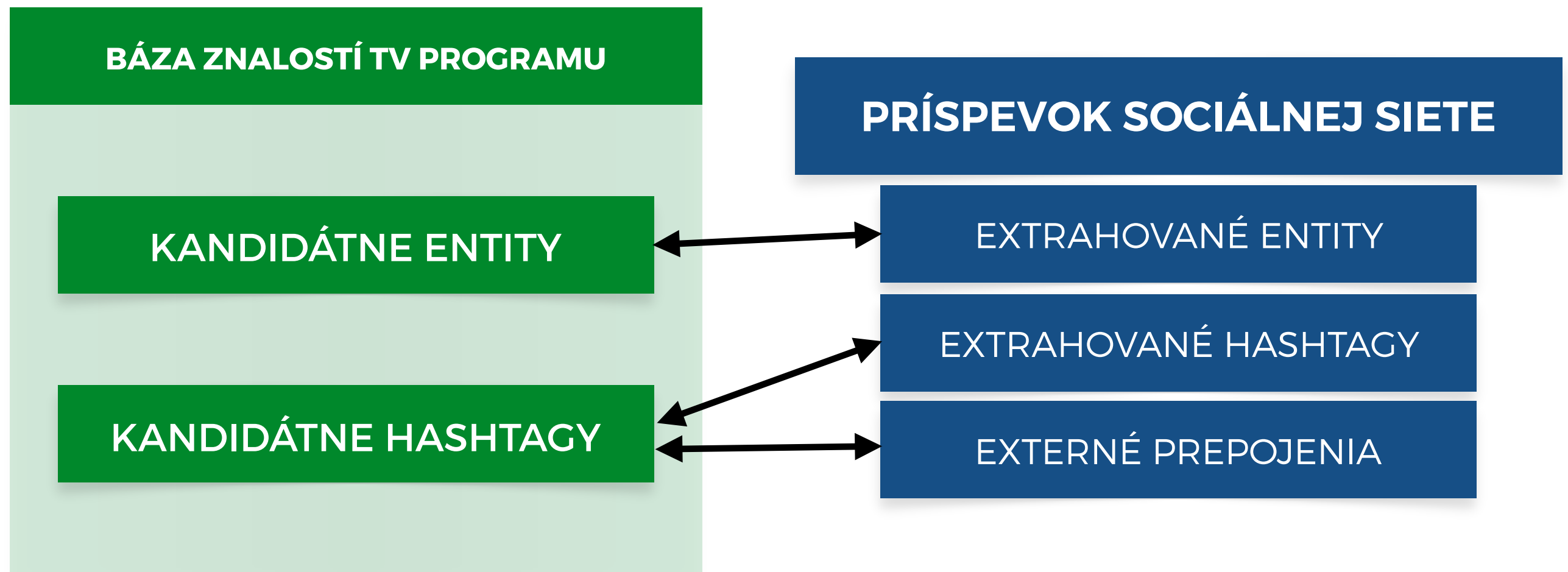
VYTVORENIE BÁZY ZNALOSTÍ



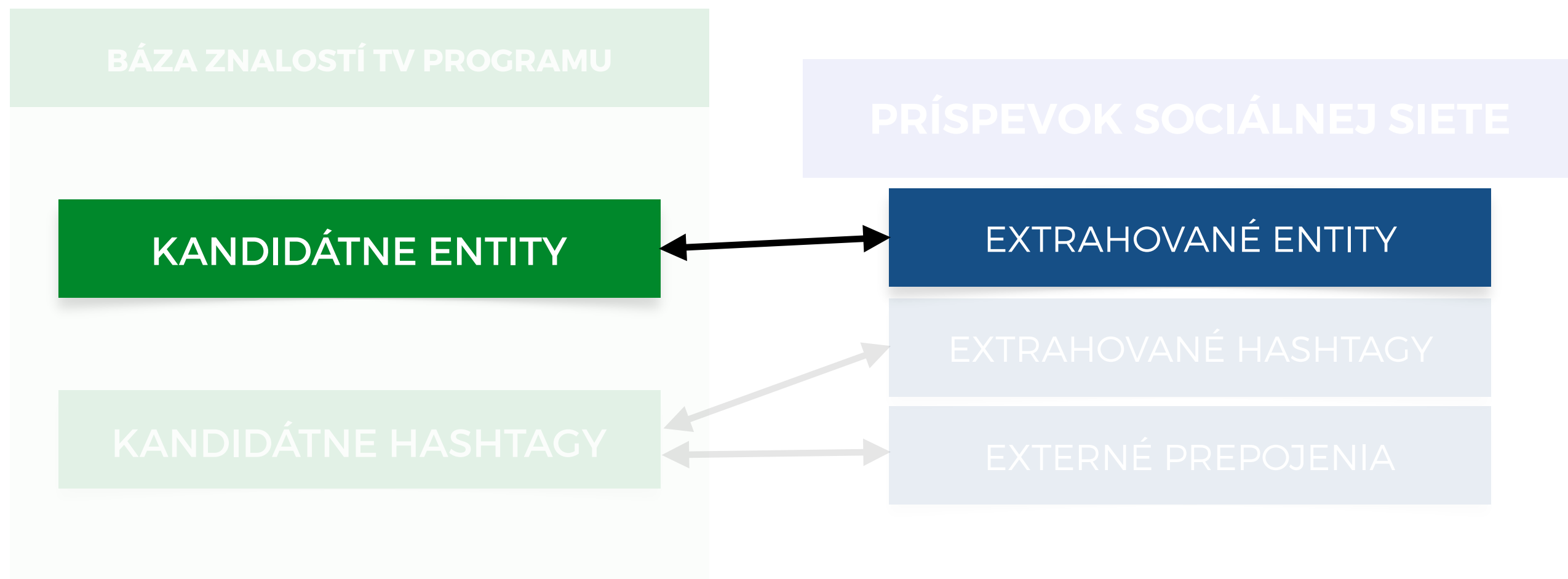
SPRACOVANIE PRÍSPEVKU SOC. SIETE



REALIZÁCIA MAPOVANIA

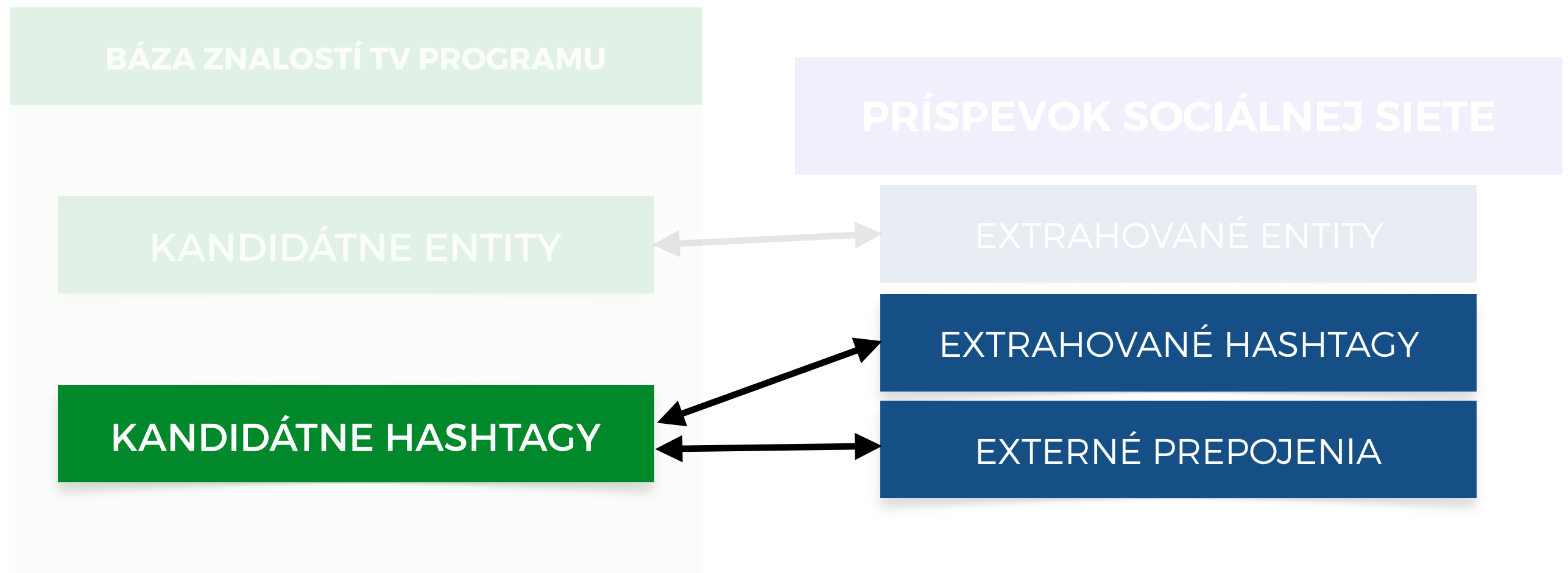


REALIZÁCIA MAPOVANIA



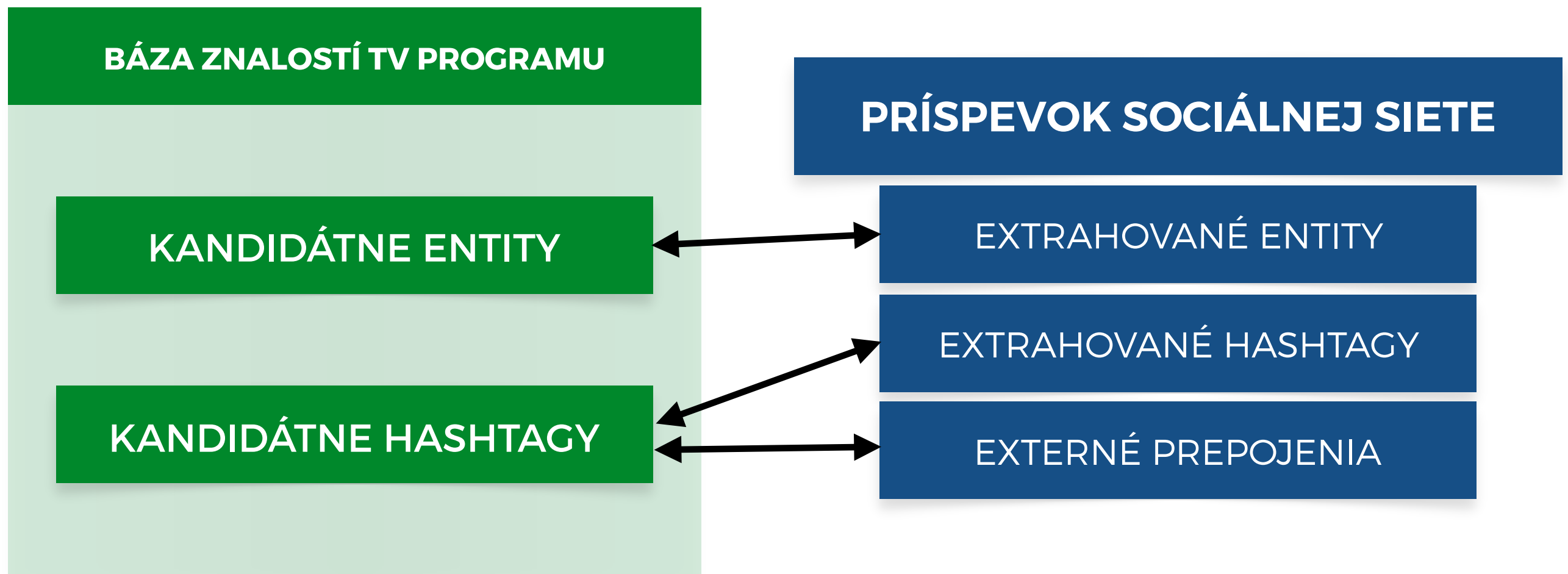
Podobnosť kandidátov:
n-gramová podobnosť
 (Kondrak, 2005)

REALIZÁCIA MAPOVANIA



Podobnosť kandidátov:
Levenshteinova vzdialenosť

REALIZÁCIA MAPOVANIA



Celková miera prepojenia medzi príspevkom a TV reláciou:
miera súvislosti

PROBLÉMY MAPOVANIA

#1

Výsledkov mapovania môže byť veľa.

Ktoré vrátiť? Ktoré prehlásiť za správne?

#2

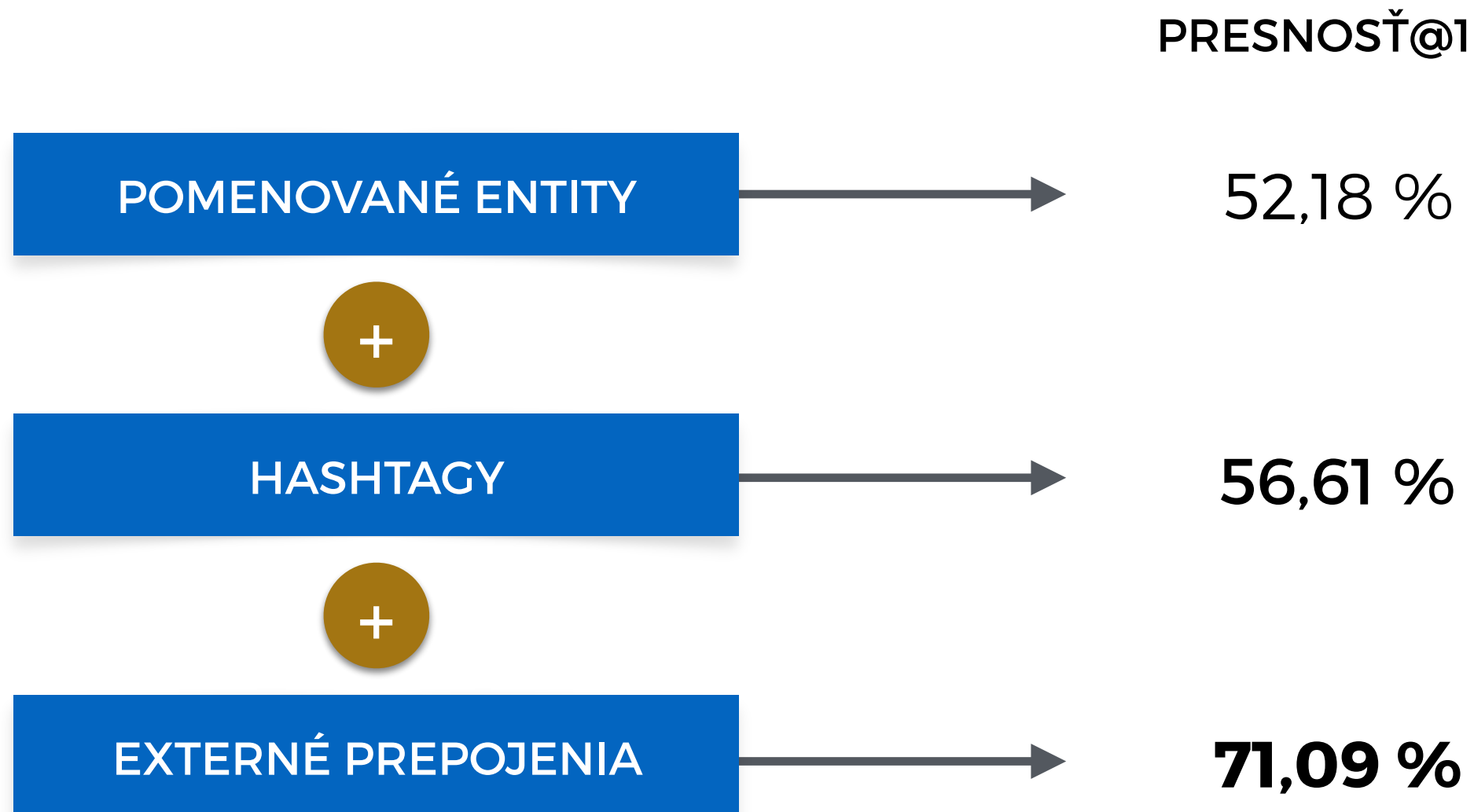
**Čo ak príspevok neobsahuje
pomenovanú entitu, hashtag a ani
externé prepojenie?**

Použitie učenia a existujúcich “bezpečných” mapovaní.

OVERENIE MAPOVANIA

- **prvá fáza:** manuálna anotácia príspevkov zo sociálnej siete
- overenie:
 - **predspracovania** (rôzne nastavenia parametrov, n-gramy alebo iný prístup),
 - **mapovania** (výsledok mapovania, stanovenie váh, rôzne nastavenie váh).
- metriky: presnosť, pokrytie, F1-skóre

POSTUPNÉ ZLEPŠENIE



Príspevky z Facebooku z obdobia: [25.12.2015 - 17.1.2016](#)

TV program z obdobia: [25.12.2015 - 17.1.2016](#).

Celkový počet príspevkov: [520](#)

VÝSLEDKY S POUŽITÍM HRANICE

Akceptovanie výsledkov mapovania, kde
miera súvislosti > hranica.

Hranica 60 až 70 %.

TV PROGRAM

HRANICA

PRESNOSŤ

POKRYTIE

-

0%

71 %

TV Kanál 1

62 %

84 %

79 %

TV Kanál 2

66 %

92 %

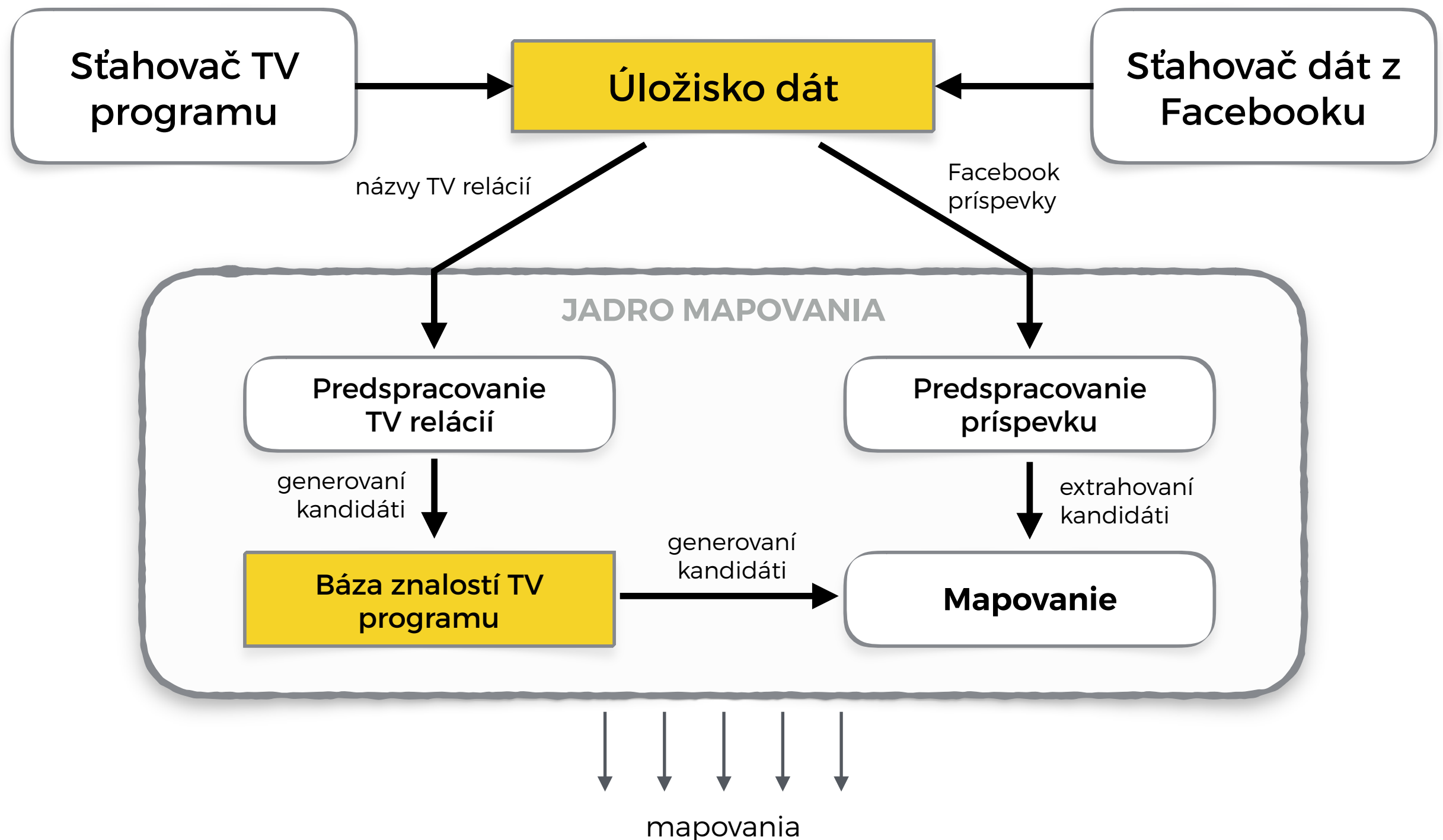
66 %

Príspevky z Facebooku z obdobia: [25.12.2015 - 17.1.2016](#)

TV program z obdobia: [25.12.2015 - 17.1.2016](#).

Celkový počet príspevkov: [520](#)

PREHL'AD RIEŠENIA



LITERATÚRA

- **Grzegorz Kondrak. 2005. N-gram similarity and distance.** In Proceedings of the 12th international conference on String Processing and Information Retrieval (SPIRE'05), Mariano Consens and Gonzalo Navarro (Eds.). Springer-Verlag, Berlin, Heidelberg, 115-126.
- **Paolo Cremonesi, Roberto Pagano, Stefano Pasquali, and Roberto Turrin. 2013. TV program detection in tweets.** In Proceedings of the 11th european conference on Interactive TV and video (EuroITV '13). ACM, New York, NY, USA, 45-54. DOI=<http://dx.doi.org/10.1145/2465958.2465960>
- **Ovidiu Dan, Junlan Feng, and Brian Davison. 2011. Filtering microblogging messages for social tv.** In Proceedings of the 20th international conference companion on World wide web (WWW '11). ACM, New York, NY, USA, 197-200. DOI=<http://dx.doi.org/10.1145/1963192.1963288>
- **Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya. 2011. Towards better TV viewing rates: exploiting crowd's media life logs over Twitter for TV rating.** In Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC '11). ACM, New York, NY, USA, , Article 39 , 10 pages. DOI=<http://dx.doi.org/10.1145/1968613.1968661>

POROVNANIE VYBRANÝCH PRÍSTUPOV

WAKAMIYA, 2011

- názvy TV programov (inv. index, Jaccardova podobnosť),
- časová a priestorová podobnosť,
- **vyžitie:** meranie sledovanosti.

CREMONESI, 2013

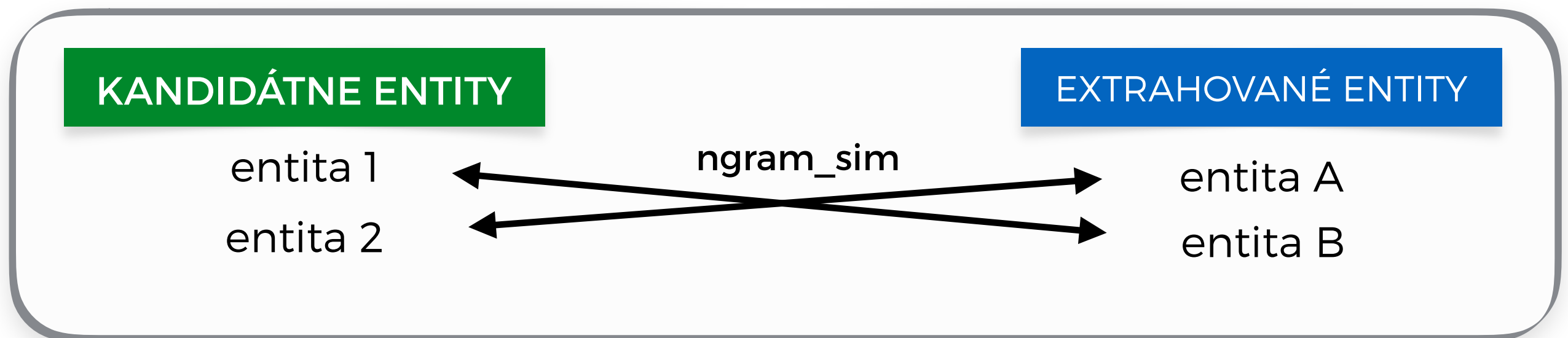
- rozpoznanie názvov TV programov v tweetoch,
- tokenizácia (vyžitie NLP), n-gramy, normalizácia cez TF-IDF
- 3 stupne: presná zhoda, voľná zhoda, klasifikácia pomocou SVM.

FENG, 2011

- názvy programov získané z IMDB.com,
- zlepšenie využitím pravidlového systému a kľúčových slov

CHARAKTERISTIKY PRÍSTUPU

- metrika podobnosti kandidátov: **n-gramová podobnosť** (Kondrak, 2005)
- n-gramy často používané v oblasti sociálnych sietí,
- slovné n-gramy vs. znakové n-gramy.

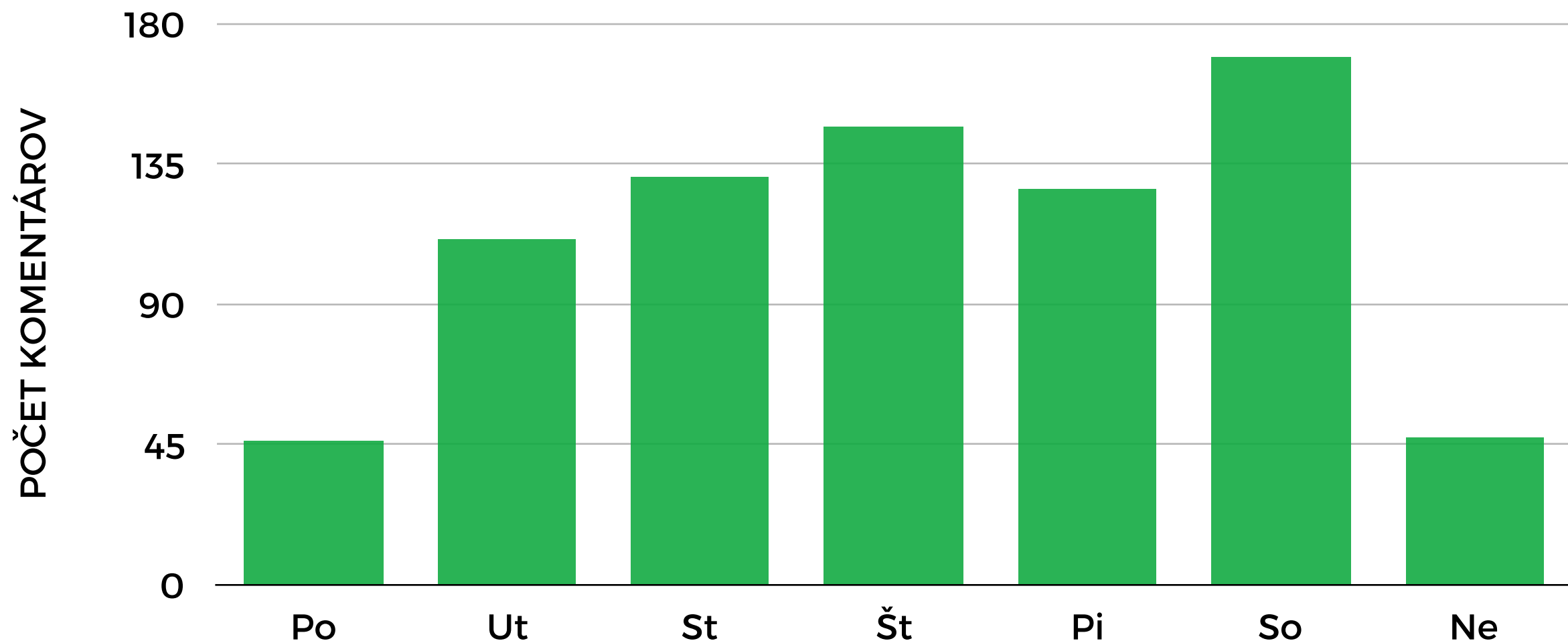


CHARAKTERISTIKY PRÍSTUPU

- **potrebné vstupy:**
 - báza znalostí - vychádza z mien TV relácií,
 - mapovanie príspevkov zo soc. siete na TV stanice,
- **výstup:**
 - miera súvislosti príspevku a TV relácie (TV relácií)

AKTIVITA NA SOCIÁLNEJ SIETI

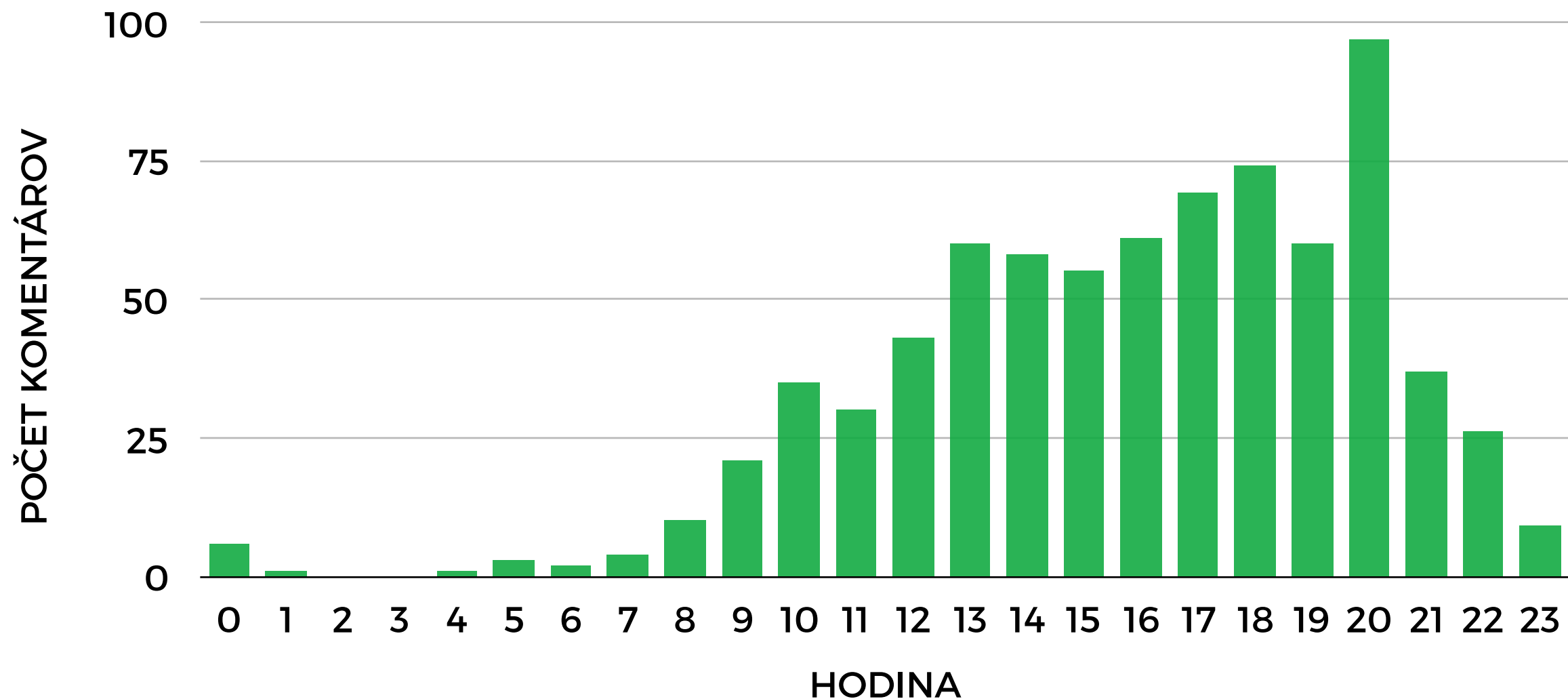
Komentáre k večernému seriálu - TV Markíza
(2.2. - 22. 2. 2015)



Vysielací čas: Ut - Št 20:30

AKTIVITA NA SOCIÁLNEJ SIETI (2)

Komentáre k večernému seriálu - TV Markíza
(2.2. - 22. 2. 2015)



Vysielací čas: Ut - Št 20:30

AKÝ TO MÁ VÝZNAM?

- napr. pre fanúšika:

