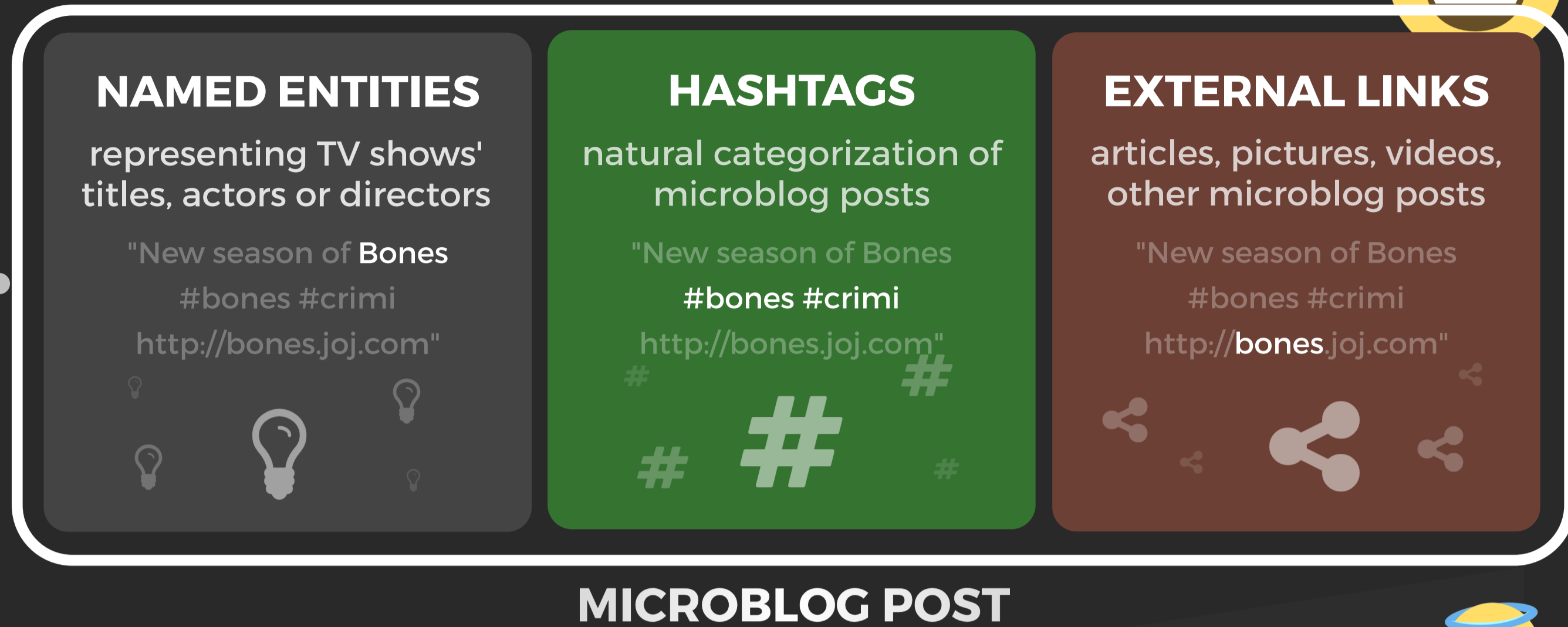
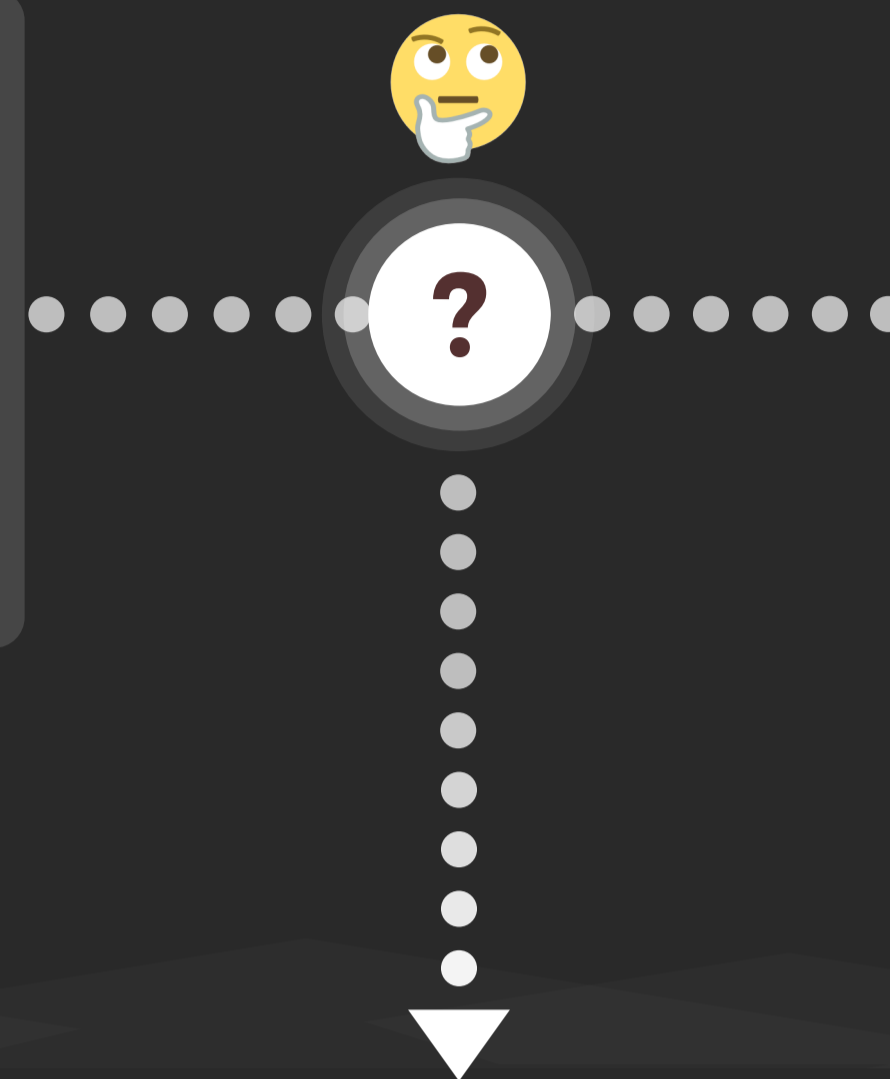


LINKING MULTIMEDIA TO MICROBLOGS FOR METADATA EXTRACTION

Peter Gašpar · Supervisor: Jakub Šimko

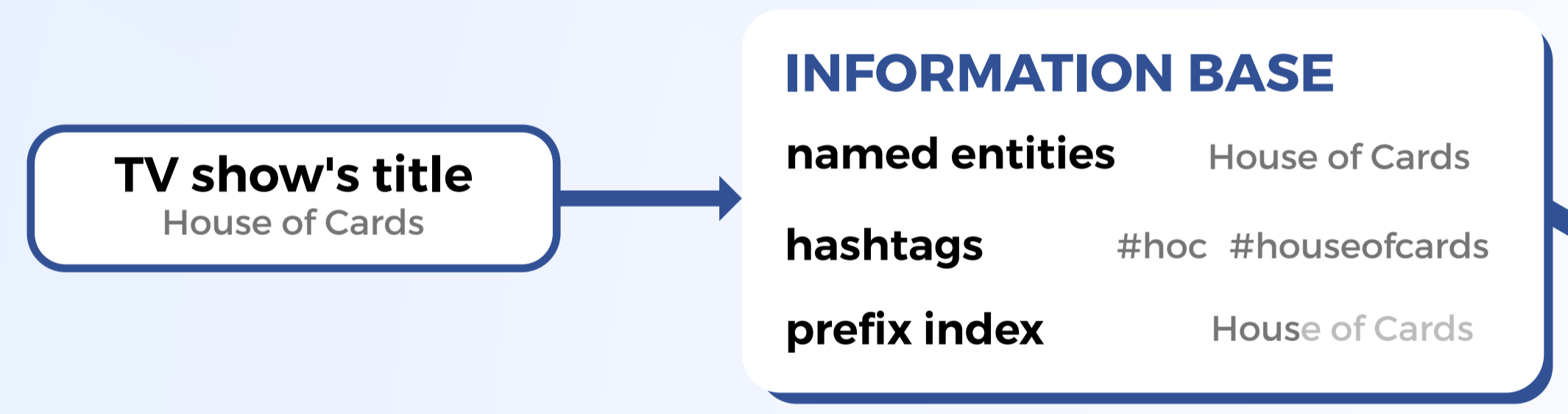
GOALS OF OUR WORK

- TV schedule contains various **TV shows** and basic information about them.
- Potentially attractive content might be hidden in **microblogs**.
- The **main goal** of our work is to find a **mapping** between a microblog post and a TV show. If we find the correct mapping, we can **extract additional metadata** that will enrich the TV schedule.

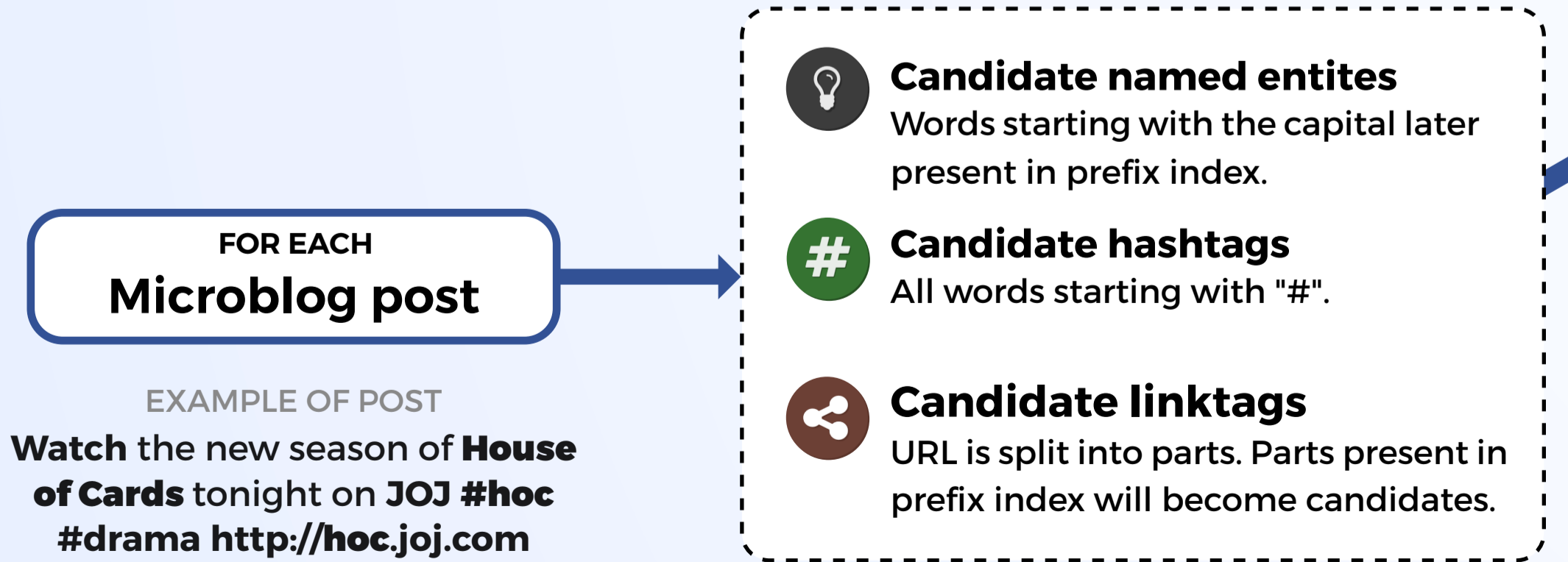


OUR MAPPING APPROACH

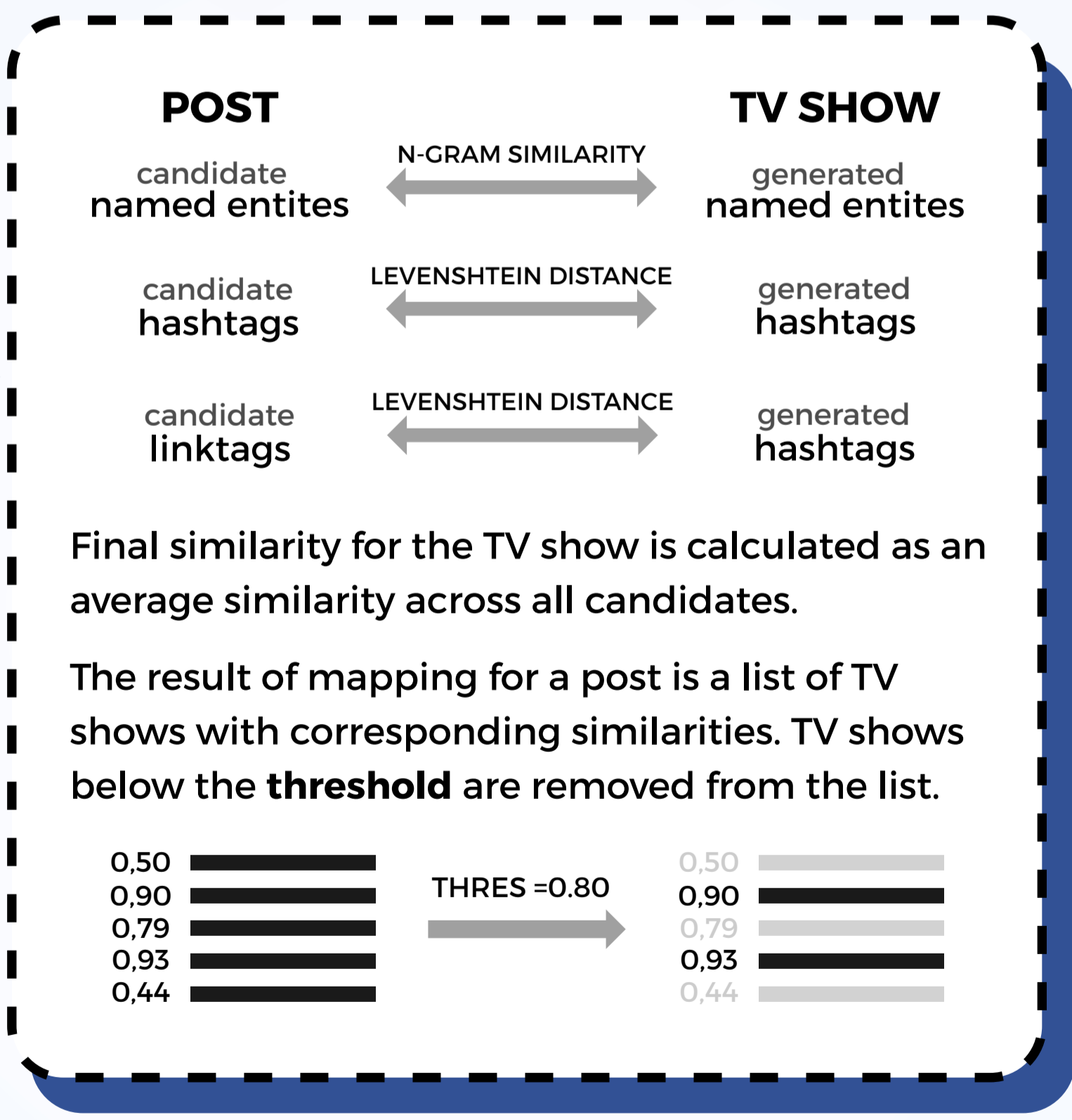
1 CREATE INFORMATION BASE
From the title of TV show we generate named entities, hashtags and prefix index. **PREFIX INDEX** is built from the first N letters of TV show's title.



2 EXTRACT FEATURES FROM POST
From each microblog post we extract three features: **named entities, hashtags, external links.**



3 FIND MAPPING
Features extracted in step 2 are searched in the information base.



4 UPDATE INFORMATION BASE
Correct mapping is used to update the information base (**co-learning**).

RESULTS

To evaluate our solution we have annotated **400 posts** from microblog (Facebook¹) of **5 Slovak** and international **TV stations**.

We improved precision by using **external links**:

PRECISION 44%	PRECISION 65%	PRECISION 86%

Using **additional features** improved precision as well:

PRECISION +7%	PRECISION +7%	PRECISION +12%
PREFIX INDEX	THRESHOLD (> 0.66)	CO-LEARNING

Full setup brought promising results:

PRECISION 86%	RECALL 83%	F1-MEASURE 85%
-------------------------	----------------------	--------------------------

Results are based on the annotated Facebook's posts published within 25.12.15 - 17.1.16 (TV Markiza, TV JOJ) and 24.3.16 - 13.4.16 (CBS, NBC, The CW).
¹ www.facebook.com