

# Structuring Wikipedia Articles with Section Recommendations

*Tiziano Piccardi, Michele Catasta, Leila Zia, Robert West*

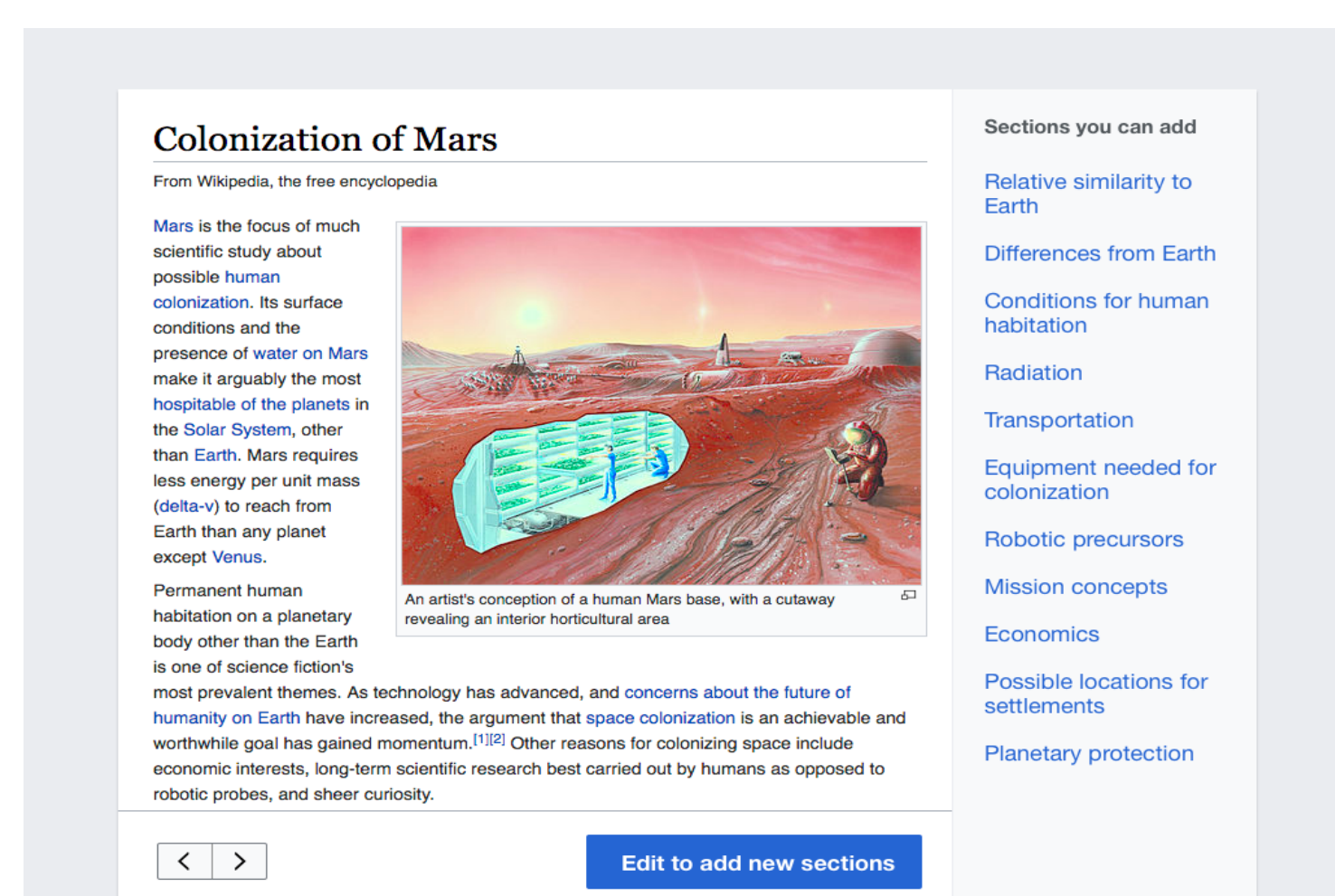
## Problem

*In the English Wikipedia alone,  
more than 2M (37%) articles  
are incomplete*

*Many articles are inconsistent  
in terms of structure  
and section titles*

*Often users are not aware of best practices  
and what sections are missing!*

## Solution



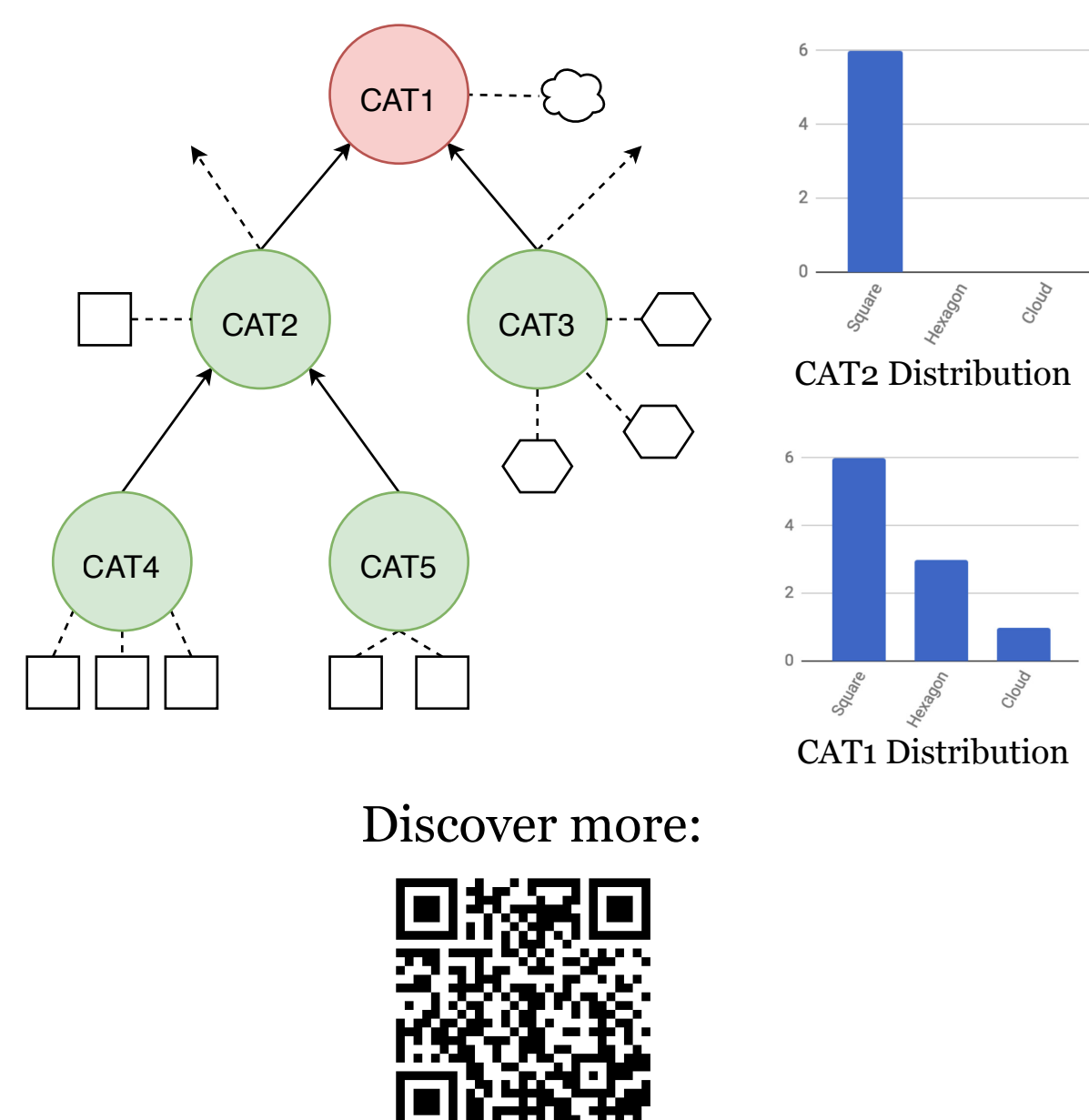
*A recommender system to guide the users  
in writing high quality articles*

*The tool must inform the editors about  
missing sections in the article*

*We can use the categories of an article to generate the list of expected sections*

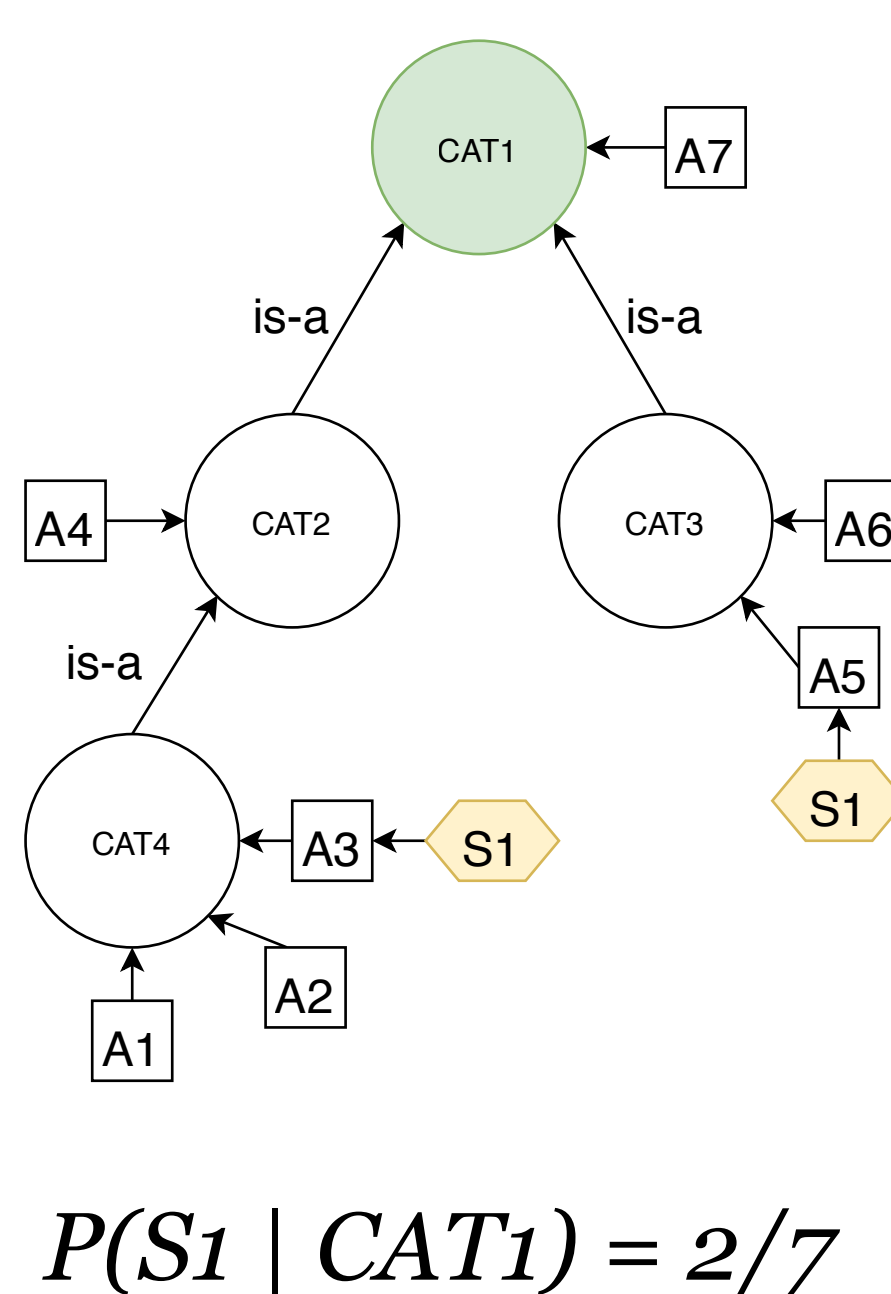
## Method

*Pruning of the category network  
according to the concept of purity*



Wikimedia Research Showcase March 2018

*Generate a taxonomy  
from the category network*



**American\_male\_film\_actors**  
Filmography: 0.59  
Career: 0.47  
Personal life: 0.38  
...

**American\_film\_producers**  
Filmography: 0.39  
Career: 0.34  
Personal life: 0.26  
...

**Living\_people**  
Career: 0.25  
Personal life: 0.17  
Biography: 0.13  
...

**Categories\_of(Leonardo DiCaprio) =**  
{American\_male\_film\_actors,  
American\_film\_producers  
Living\_people, ...}

$\sum_{ij} (P(S_i | C_j))$   
 $C_j \in \text{Categories\_of(Leonardo DiCaprio)}$

**Leonardo DiCaprio**

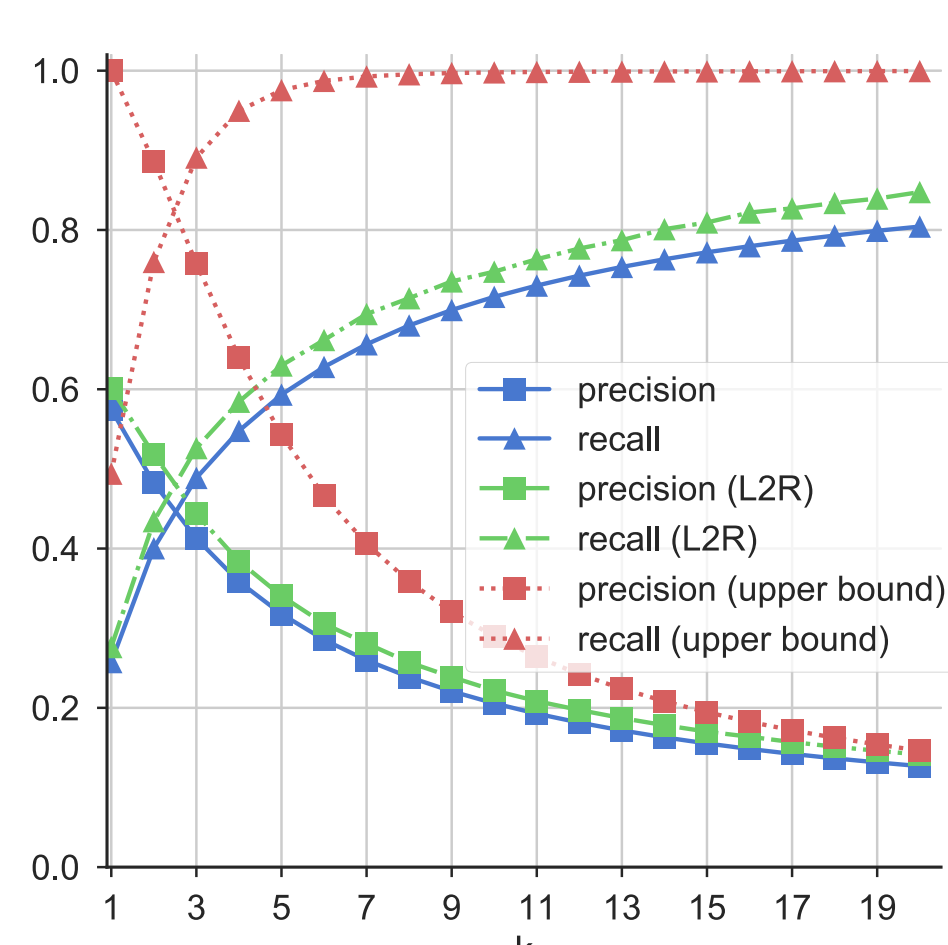
Career: 9.98  
Filmography: 9.51  
Personal life: 9.48  
Early life: 8.40  
Awards and nominations: 2.45

*Probability  $P(S | C)$  of observing  
section  $S$  in category  $C$   
(and its subcategories)*

*Merging of the recommendations  
generated by the categories  
of the article*

## Results

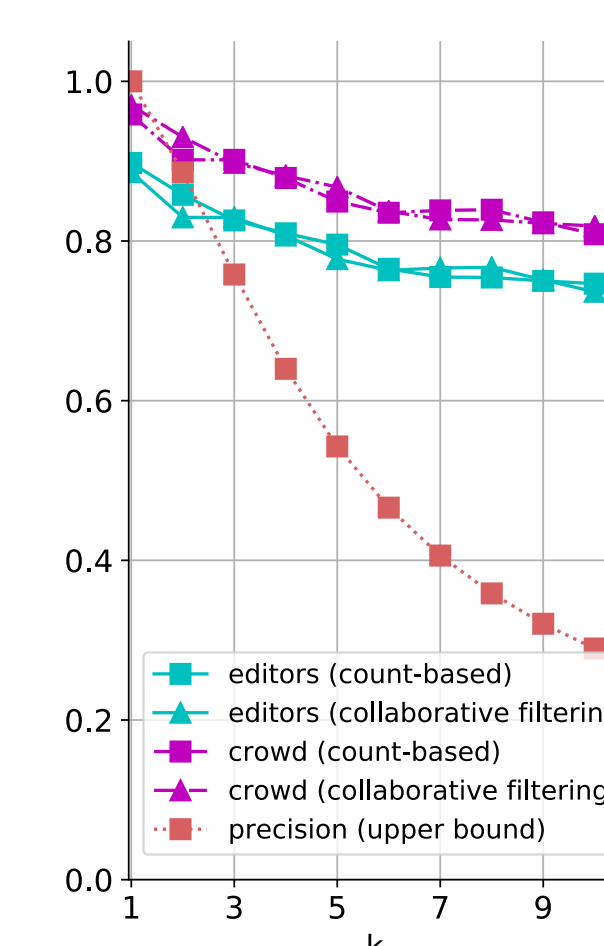
### Automatic evaluation:



*Precision@10 = 20%*  
(max possible 28%)

*Recall@10 = 72%*  
(max possible 98%)

### Human evaluation:



*Expert editors:*  
*Precision@10 = 72%*

*Crowd-workers:*  
*Precision@10 = 81%*

*Future work: Deploy the system in Wikipedia!*

