



Measuring Patent Similarity Based on SAO Semantic Analysis

Yun Chen¹ · Xuefeng Wang¹ · Huichao Ren¹ · Ying Wang¹ · Zhinan Wang¹

Xuefeng Wang

[Email: wxf5122@bit.edu.cn](mailto:wxf5122@bit.edu.cn)

1. School of Management and Economics, Beijing Institute of Technology

Yun Chen

Email: chenyun_helloworld@foxmail.com


KMDA

Contents

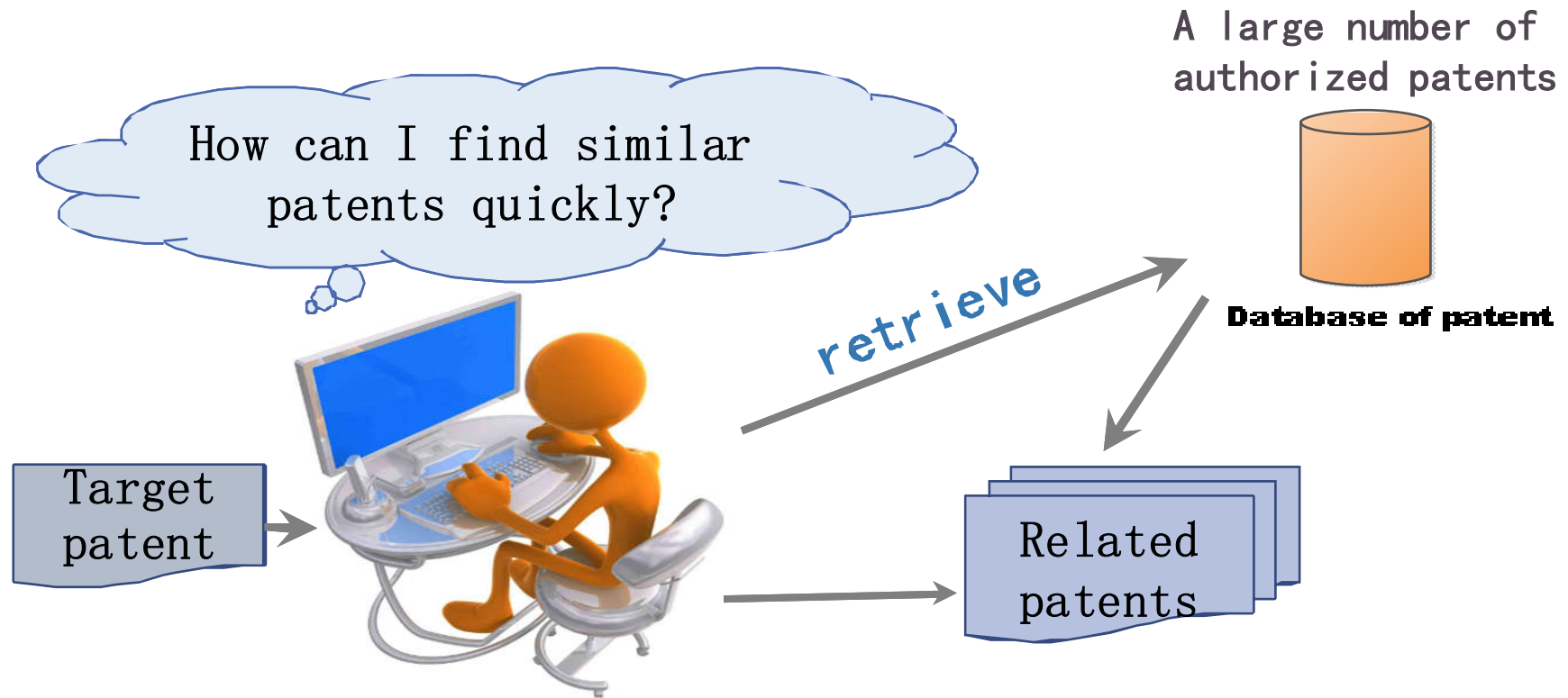
- Background
- Introduction
- Methodology
- Case Study
- Conclusion




Background

- Technology is a significant factor for development of companies and countries.
 - Patent is the most important pattern to protect the intellectual properties of technology.
 - Applicants and examiners have the requirement to search and find similar patents.
- 

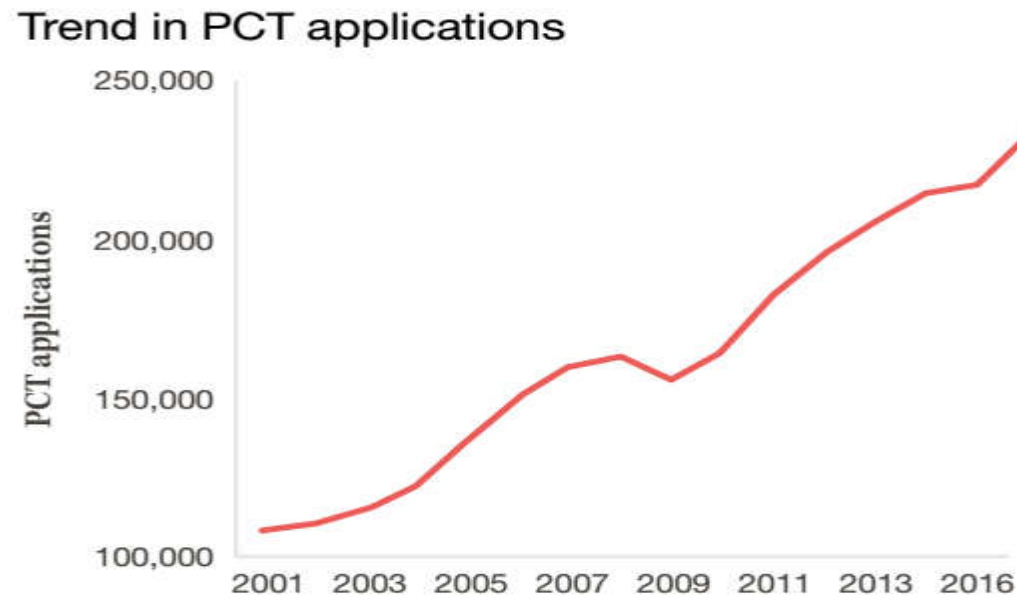
Background



Background

- Technology is a significant factor for development of companies and countries.
 - Patent is the most important pattern to protect the intellectual properties of technology.
 - Inventors, applicants and examiners have the requirement to search and find patents.
 - The patent quantity is increasing rapidly.
- 

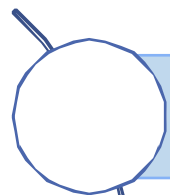
Background



Patent Cooperation Treaty Yearly Review 2017 —WIPO

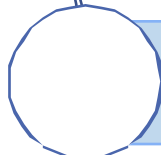
Introduction

Three mainly methods to measure patent similarity



IPC code analysis

Vague classification.



Citation analysis

Not all database provide citation information. And it doesn't work well for new patents.



Keywords-based analysis

Cannot express the semantic technology information well.

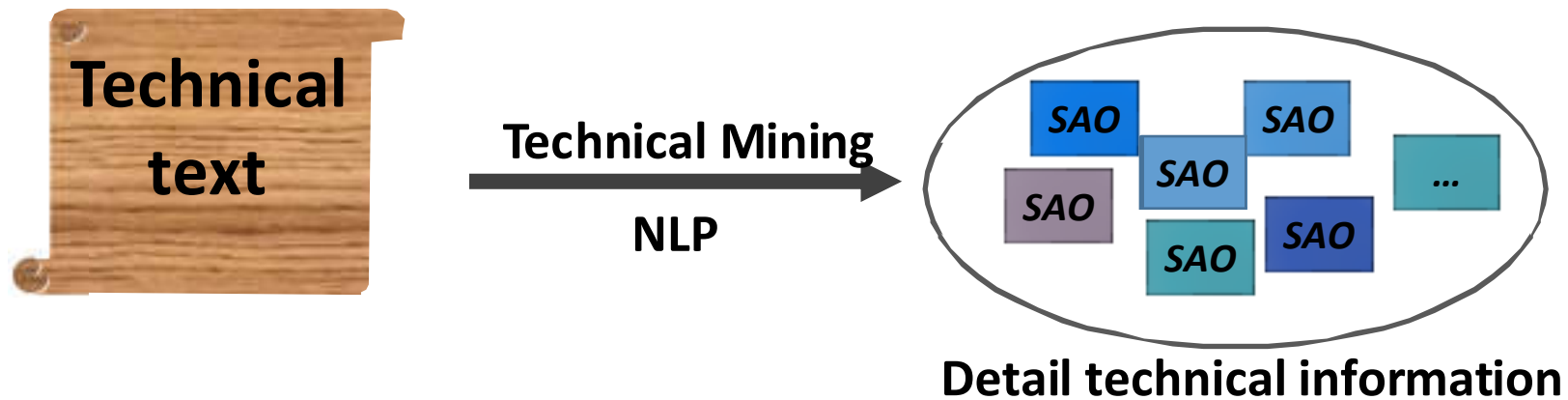


Introduction

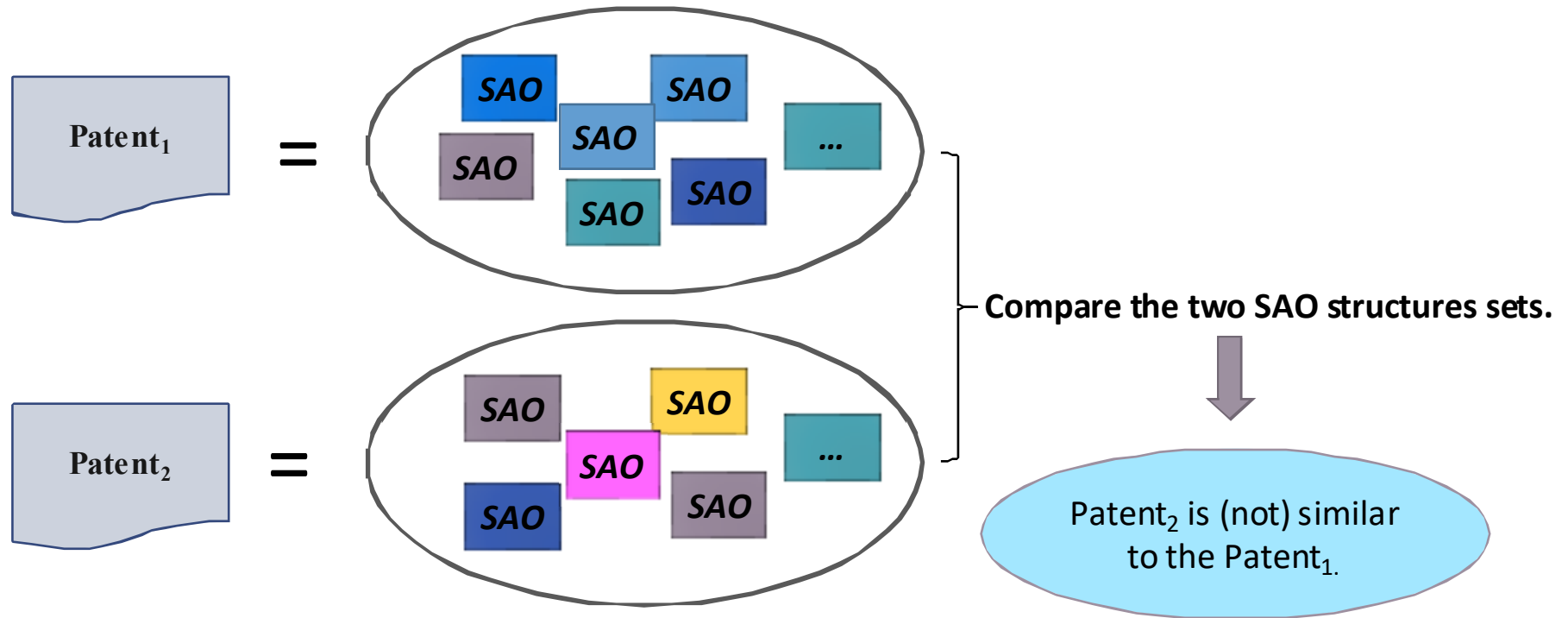
SAO(Subject-Action-Object) structure analysis is a hotspot. SAO structures are composed of Subject (noun phrase), Action (verb phrase) and Object (noun phrase). It emphasizes the “key concepts” and can provide various technology information on their semantic relationships.

Eg. Battery energizes bulb. → Battery energizes bulb

S
A
O

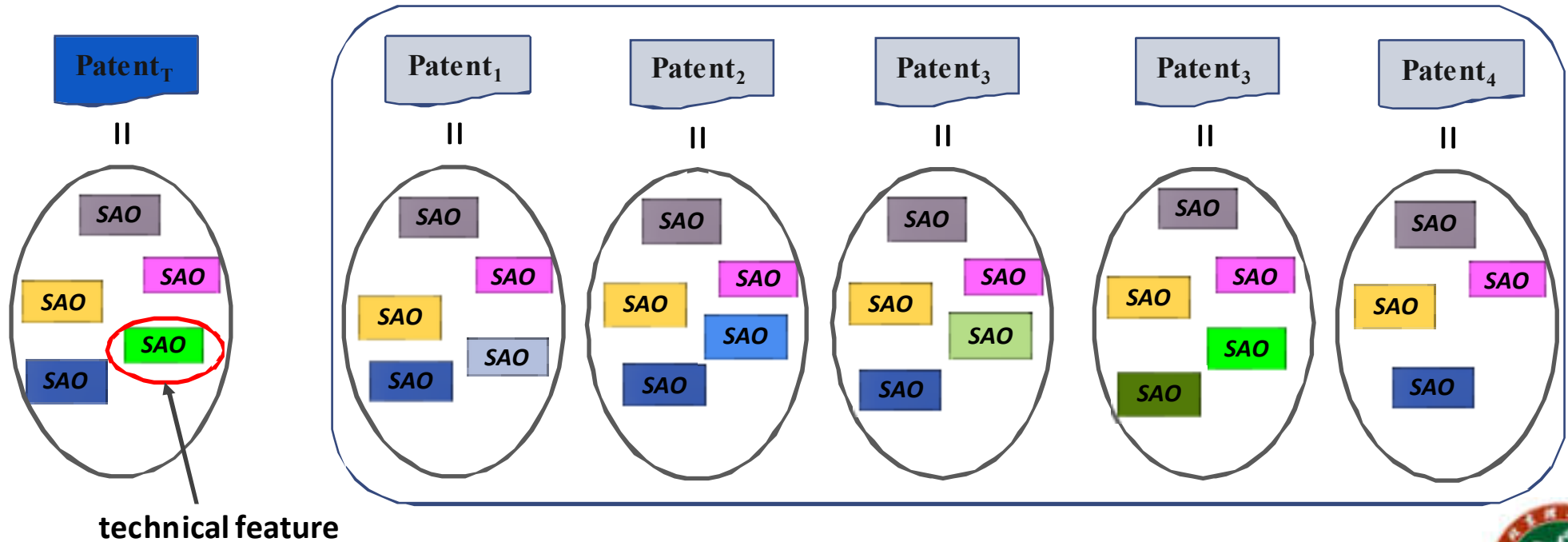


Introduction



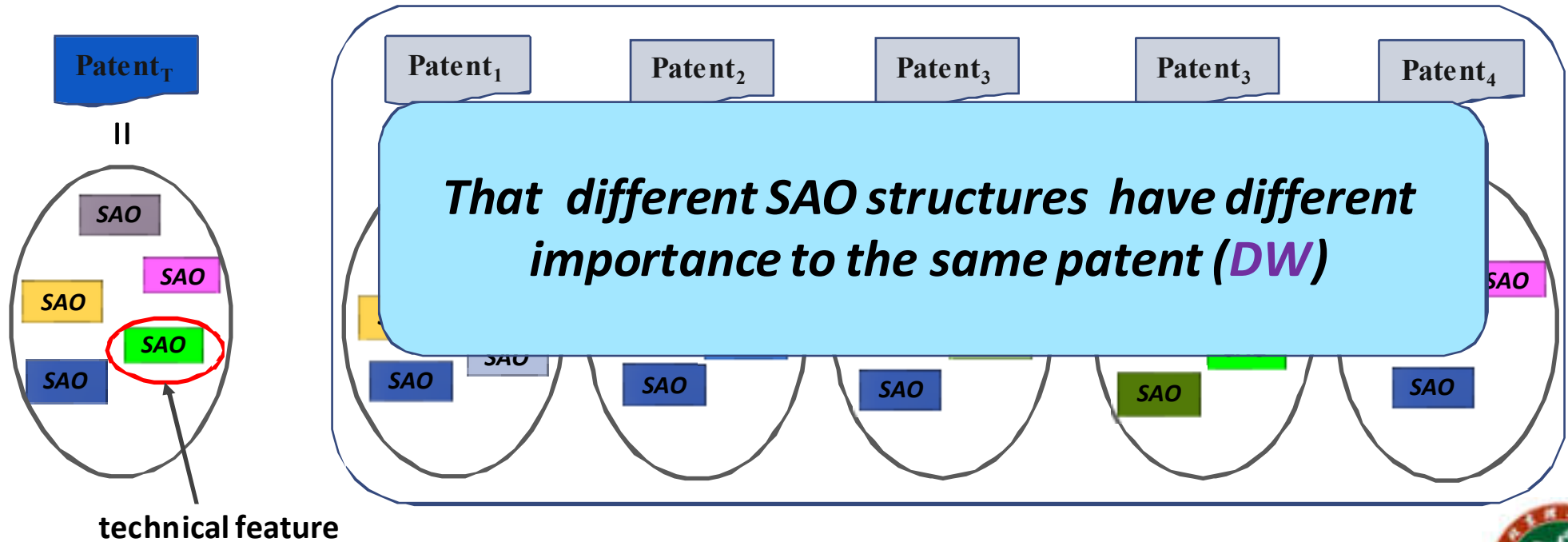
Methodology

The question: Find the most similar patent to the Patent_T from Patent₁ to Patent₄



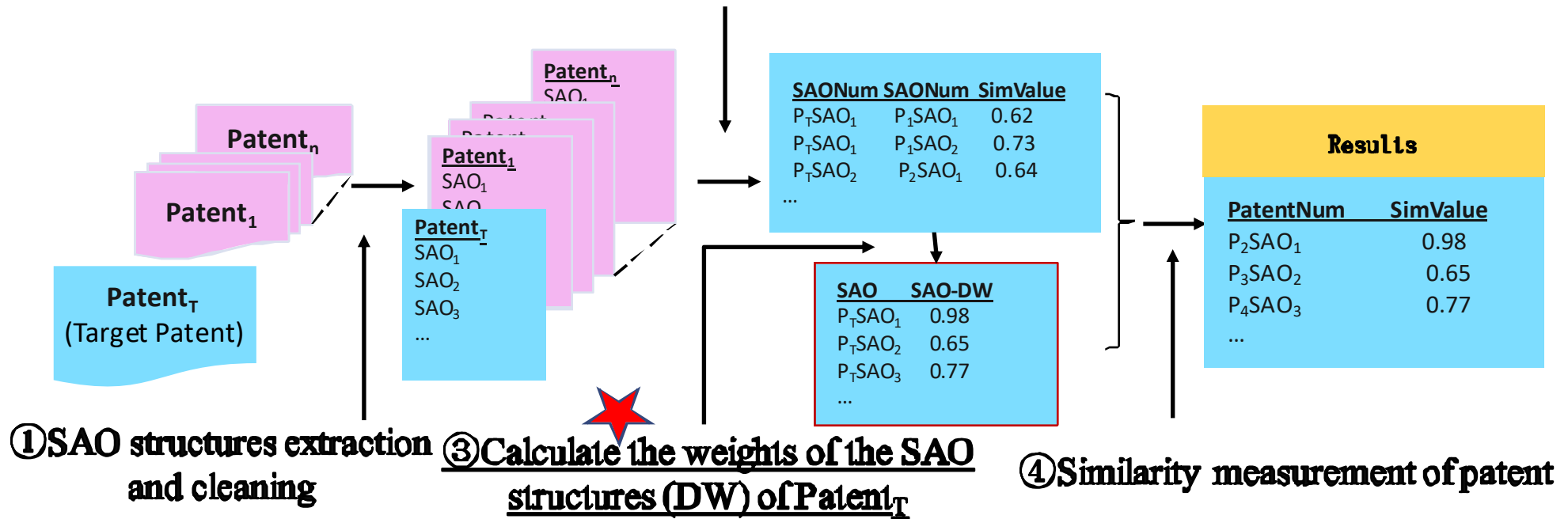
Methodology

The question: Find the most similar patent to the Patent_T from Patent₁ to Patent₄



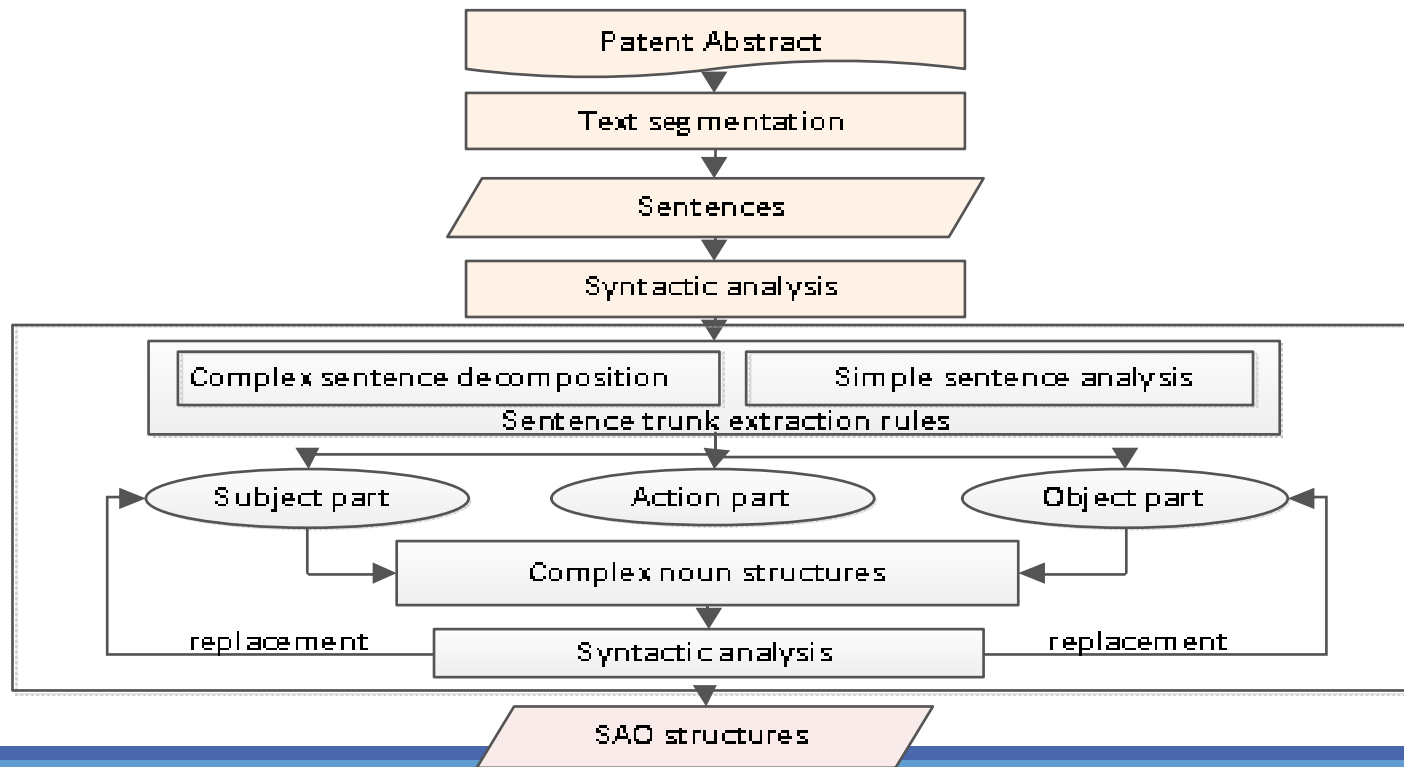
Methodology

② SAO structure semantic similarity measurement based on WordNet



Methodology

① SAO structures extraction and cleaning



Methodology

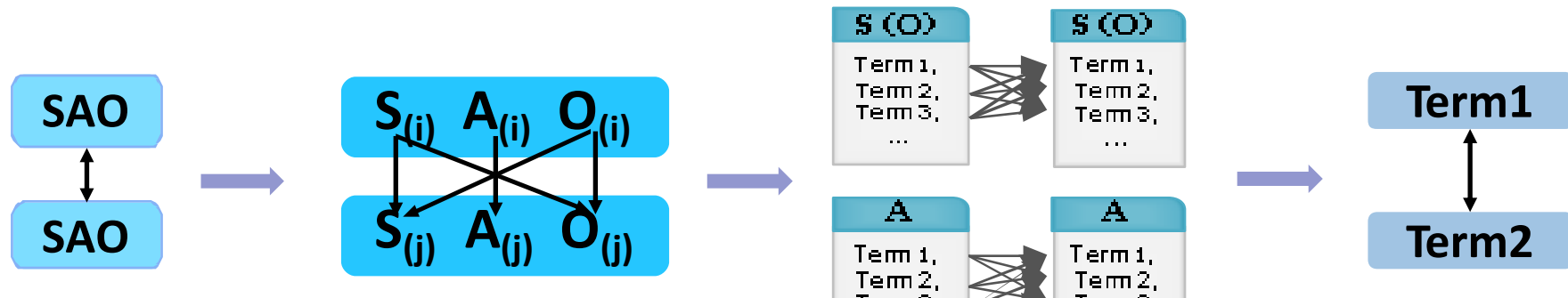
① SAO structures extraction and cleaning

	Steps
1	Split the S and O components of SAO structures that are too long
2	Remove the meaningless SAO structures
3	Remove the stop words
4	Change the abbreviations
5	Word part of speech reduction



Methodology

② SAO structure semantic similarity measurement based on WordNet



$$Sim(SAO_i, SAO_j) = \begin{cases} \alpha * \frac{[Sim(S_{(i)}, S_{(j)}) + Sim(O_{(i)}, O_{(j)})]}{2} + \beta Sim(A_{(i)}, A_{(j)}), & \textcircled{1} \\ \alpha * \frac{[Sim(S_{(i)}, O_{(j)}) + Sim(O_{(i)}, S_{(j)})]}{2} + \beta Sim(A_{(i)}, A_{(j)}), & \textcircled{2} \end{cases}$$

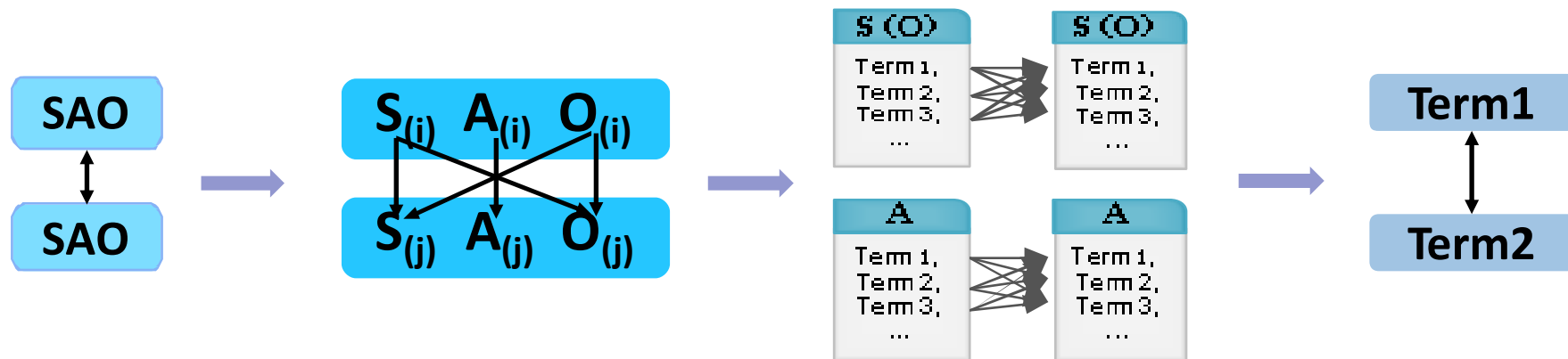
① $Sim(S_{(i)}, S_{(j)}) + Sim(O_{(i)}, O_{(j)}) \geq Sim(S_{(i)}, O_{(j)}) + Sim(O_{(i)}, S_{(j)})$ set1 and synset2

② $Sim(S_{(i)}, S_{(j)}) + Sim(O_{(i)}, O_{(j)}) < Sim(S_{(i)}, O_{(j)}) + Sim(O_{(i)}, S_{(j)})$ coefficient: α, β



Methodology

② SAO structure semantic similarity measurement based on WordNet



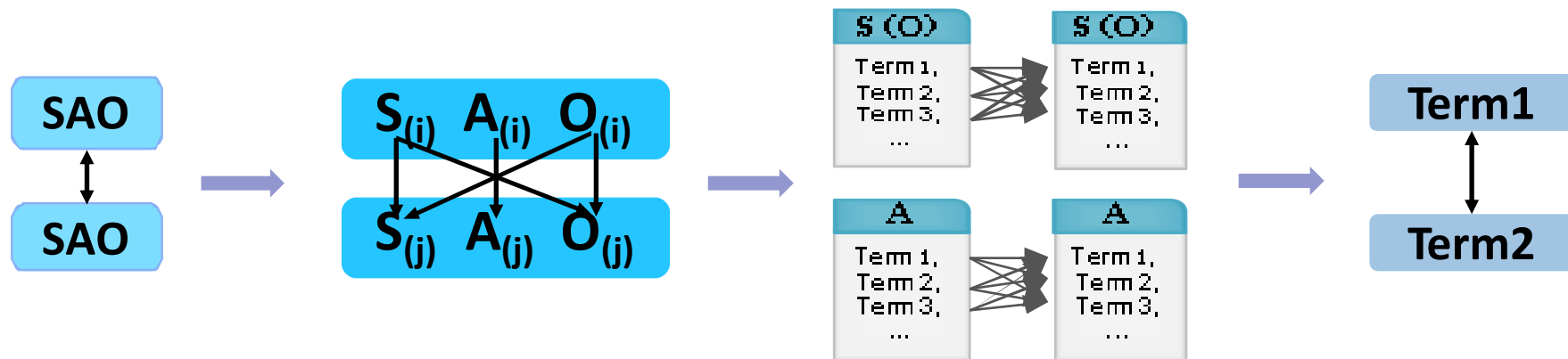
$$Sim(N_i, N_j) = \frac{2 * Match(N_i, N_j)}{NumTermN_i + NumTermN_j}$$

N is S(A or O) of the SAO structure.



Methodology

② SAO structure semantic similarity measurement based on WordNet



$$Sim(Term1, Term2) = \frac{2 * IC(Lcs)}{IC(Term1) + IC(Term2)}$$

lcs = the least common subsumer of synset1 and synset2

IC = the Information Content (of a synset).



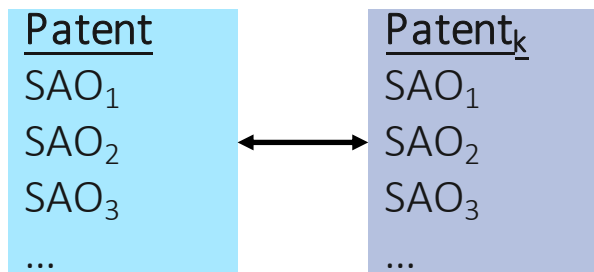
Methodology

- ③ Calculate DW which is the weight of the SAO structure of Patent_T

$$DW_{SAO_i} = 1 - \frac{F}{N + 1}$$

N: N is the number of the related patents.
F: F is the number of the related patents that contain the SAO_i .

- ④ Similarity measurement of patent



$$Sim(P, P_k) = \frac{2 * \sum_{i=1}^m DW_{SAO_i} * Match_{SAO_i}}{Num_{SAO_P} + Num_{SAO_{P_k}}}$$

$$Match_{SAO_i} = \begin{cases} 0, & SAO_i \text{ is appearance in Patent}_k \\ 1, & SAO_i \text{ isn't appearance in Patent}_k \end{cases}$$



Case Study

The technology topic : the robot docking station.

Database: Derwent Innovation

Date: 1997-2017.06.20

Patent retrieval query	Record
((ALLD =(ROBOT) OR ALLD =(ROBOTICS) OR ALLD =(ROBOTIC)) OR (AIC=(B25J000916) OR AIC=(B25J OR 000920) OR AIC=(B25J00090003) OR AIC=(B25J00110005) OR AIC=(B25J00110015) OR AIC=(B60W0030) OR AIC=(B60W2030) OR AIC=(Y10S000901) OR AIC=(G05D00010088) OR AIC=(G05D000102) OR AIC=(G05D000103) OR AIC=(G05D22010207) OR AIC=(G05D22010212))) AND (ALLD=(Dock* station)))	220



Case Study

The information of the target patent with the latest publication date. We calculated the similarity between the target patent and every patent in the 220 retrieved patents. The 220 patents are numbered 1 to 220.

Publication Number	FR3046259A1
Title-DWPI	Docking station for mobile robot, has set of infrared LEDs arranged to emit attracting rays in robot approach region, and another set of infrared LEDs arranged to emit repelling rays outside robot approach region
Abstract - DWPI	The docking station (10) has a set of infrared LEDs (21-23) arranged around a robot parking zone, so as to emit attracting rays (R1-R3) in a robot approach region. Another set of infrared LEDs (24, 25) are arranged on each side of the robot parking zone, so as to emit repelling rays (R4, R5) outside the robot approach region, where the repelling rays have a shorter range than the attracting rays. The former set of LEDs is arranged such that the attracting rays are emitted in directions (X1-X3) intersecting at a fixed point (P) of the robot parking area. Docking station for a mobile robot. The sets of infrared LEDs emit attracting rays and the repelling rays, respectively, thus ensuring a mobile robot to approach the docking station according to appropriate directions defined by the attracting rays while avoiding approach to the docking station in improper directions defined by the repelling rays. The drawing shows a schematic top view of a docking station showing attracting rays and repelling rays. PFixed pointR1-R3Attracting raysR4, R5Repelling rays X1-X3 Directions 10 Docking station 21-25 Infrared LEDs.
Publication Date	2017/6/30



Case Study

2,833 SAO structures were extracted from abstracts of the 220 patents. 2,744 SAO structures are obtained when we accomplished the cleaning progress of SAO structures.

SAO structure examples

S(Subject)	A(Action)	O(Object)
attract ray	in	robot approach region
emit repel ray	outside	robot approach region
avoid approach	in	improper direction
improper direction	define	repel ray
schematic top view	of	docking station
docking station	attract	ray



Case Study

The examples of similarity between SAO structures

NO. P_T	NO. SAO of P_T	NO. Patent	NO. SAO of Patent	Similarity between SAO structures
1	4	2	13	0.400
1	6	2	1	0.400
1	7	2	1	0.360
1	8	2	1	0.400
1	11	2	1	0.400
1	14	2	1	0.400
1	14	2	13	0.520
1	15	2	1	0.400
1	1	3	15	0.200
1	2	3	15	0.200
1	3	3	15	0.360
1	4	3	16	0.400
1	5	3	1	0.200
1	6	3	1	0.600
1	6	3	2	0.200



Case Study

The DWs of each SAO structure of Patent_T

No. SAO	DWSAO	No. SAO	DWSAO
1	0.9591	9	0.8227
2	0.95	10	0.8318
3	0.8045	11	0.7136
4	0.9636	12	0.9818
5	0.8773	13	0.8409
6	0.65	14	0.5409
7	0.6545	15	0.4045
8	0.6091		



Case Study

Patent similarity values:

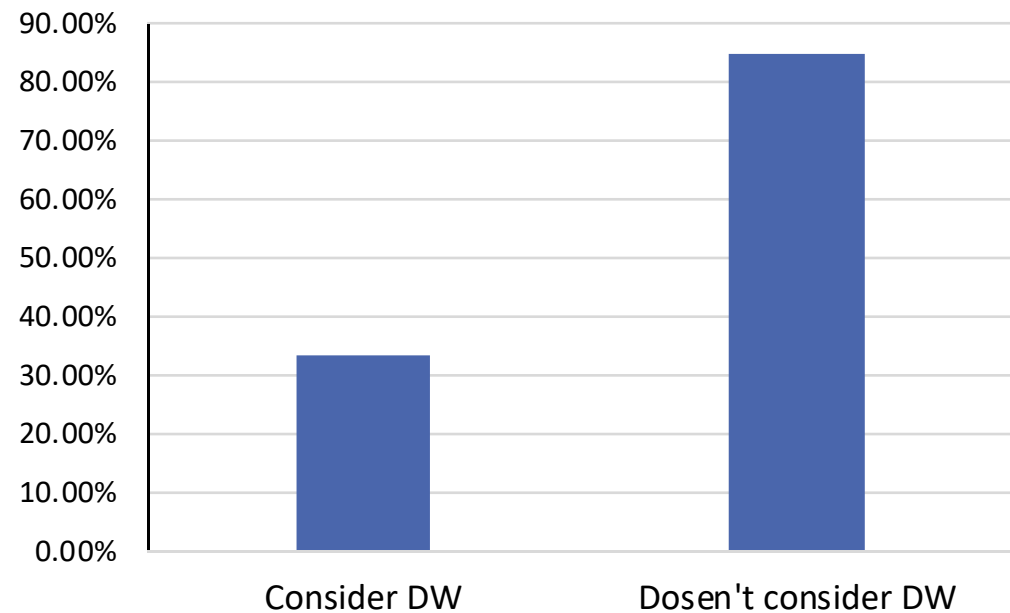
Some patent similarity values between related patent and Patent _T without considering the DW			
NO. Patent	similarity	NO. Patent	similarity
1	1.000	81	0.385
77	0.500	104	0.381
8	0.476	97	0.370
76	0.438	133	0.370
5	0.429	7	0.357
90	0.387	44	0.357

Some patent similarity value between related patent and Patent _T considering the DW			
NO. Patent	similarity	NO. Patent	similarity
1	1.000	38	0.346
77	0.482	5	0.344
8	0.400	104	0.328
76	0.368	44	0.324
212	0.357	162	0.313
166	0.351	133	0.312



Case Study

The proportion of duplicate value of patent similarity in the total number of patents (220).



Case Study

- Select 13 patents
- read the abstracts and sort them (Rank)
- Rank1 is the rank of 13 patents by patent similarity values that were calculated basing on general SAO semantic analysis
- Rank2 is rank of 13 patents by patent similarity values that were calculated basing on our SAO semantic analysis considering the different weight of the SAO structures.

The different sorting comparison			
No. Patent	Rank	Rank 1	Rank 2
77	1	3	3
168	2	8	5
215	3	7	4
94	4	12	12
8	5	2	2
14	6	11	10
96	7	10	11
5	8	4	6
91	9	5	8
109	10	9	9
105	11	6	7
34	12	13	13
78	13	1	1
Rank change value sum	—	58	46
Average ranking change	—	4.461538	3.538462



Conclusion

The case study to measure the similarities of patents about robot technology demonstrates the reliability of our method and the results indicate the practical meaning of our method to get more accurate result than previous methods.

In the future we can consider using better semantic similarity calculation method and using machine learning to improve the accuracy.





Thank you !

Yun Chen

Email: chenyun_helloworld@foxmail.com

KMDA

A solid blue horizontal bar spans the width of the slide at the bottom.