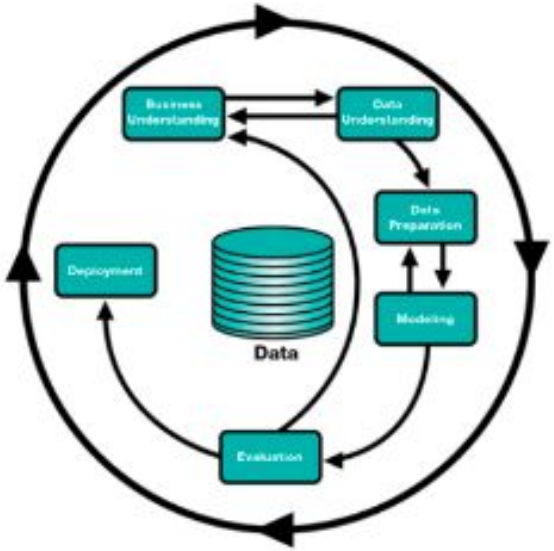
Patent Recommendation System

TEAM #7
Aaron Lee, Hongfei Xu
Juan Chen, Xiaoting Jin

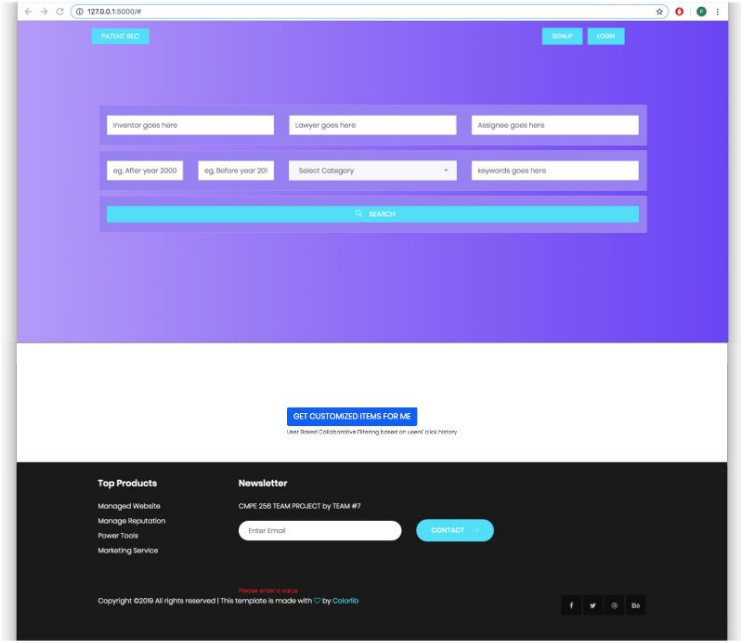
Work Before Mid-term Review

- Data Collection and Data Processing
  - Patents from www.uspto.com and www.patentsview.org
  - Feature extraction
    - CB: Title and Abstract
    - KB: Attributes (section, date, inventor, CPC classification, etc)
  - Cleaning data
- System and User Interface Design
- Individual Task Allocation



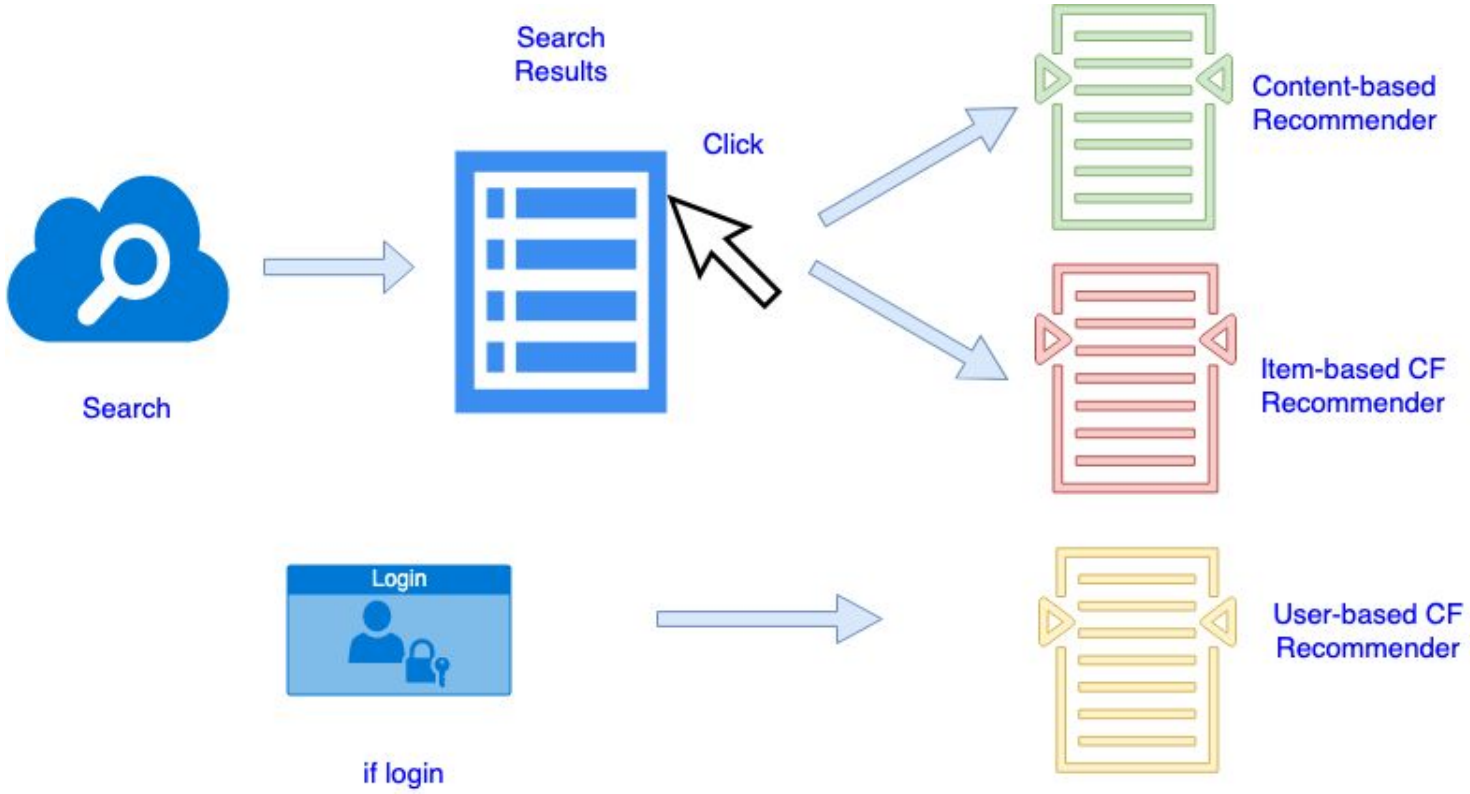
Work After Mid-term Review

- System Implementation
  - Hybrid pipelined system (CB, KB)
  - Collaborative Filtering (User-Based, Item Based)
- UI Testing (Web App)
- Presentation
  - Review of Implementation and Results
  - Project Evaluation



Web application preview

System Design

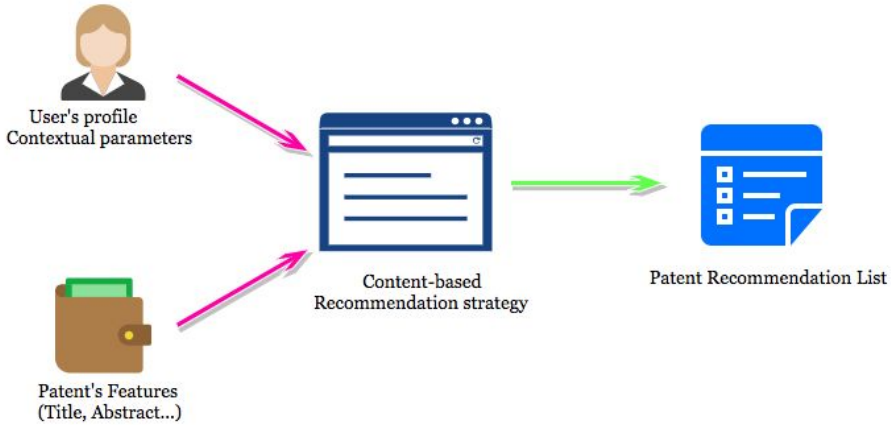


Implementation : Content-based Method

- Input features: context of titles and abstracts (string objects)
- Output: A list of recommending patents

	id	title
0	10000000	Coherent LADAR using intra-pixel quadrature de...
1	10001015	Drag reduction systems having fractal geometry...
2	10002022	Processing interrupt requests
3	10003026	Ladder tetrazine polymers
4	10004044	Communication control apparatus and wireless c...

	id	abstract
0	10000000	A frequency modulated (coherent) laser detecti...
1	10001015	Airfoil and hydrofoil systems include structur...
2	10002022	A method, a computer program product, and a co...
3	10003026	A ladder tetrazine polymer is disclosed.
4	10004044	[Object] To achieve both prevention of harmful...



Implementation : Content-based Method (Continued)

- Recommending from context keywords
- Algorithm: TF-IDF
- Implementation steps:
  - Context Pre-processing (removing punctuations, stop words and stemming…)
  - Calculating TF-IDF scores for context in title and abstract by using *TfidfVectorizer*
  - Obtaining sum of scores from above features with different weights (80% for title and 20% for abstract)

aa	aad	abbe	abc	abdominal	aberrations	ablatable	ablation	abnormalities	aborting	...
0.0	0.0	0.0	0.0	0.729313	0.0	0.0	0.0	0.0	0.0	...
0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...
0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...
0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...
0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...

Overview of TF-IDF Vectors

	id	result_a	result_t	result_weighted
920	4621435	0.286971	0.0	0.057394
4058	7776902	0.221397	0.0	0.044279
4385	8105180	0.185603	0.0	0.037121
785	4486296	0.166784	0.0	0.033357
3675	7392197	0.150608	0.0	0.030122

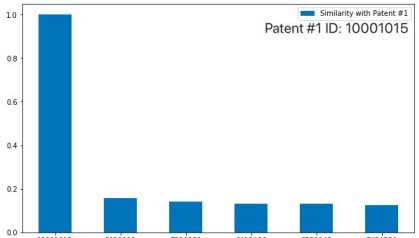
Scores of recommendation with keyword “able”

Implementation : Content-based Method (Continued)

- Recommending patents with similar context features
- Algorithm: TF-IDF and Cosine Similarity
- Implementation steps:
  - Calculating cosine similarities between selected patent and the rest from TF-IDF vectors

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

  - Ranking the patents in descending order of cosine similarity and returning the top 5 similar patents



Top 5 recommendation for patent #1 (cosine similarity)

	sim_temp_t	sim_temp_a	Similarity with Patent #1
1	1.000000	1.000000	1.000000
1320	0.192589	0.019586	0.157988
3809	0.173271	0.010644	0.140745
2700	0.162712	0.010217	0.132213
2839	0.160076	0.016350	0.131331
3737	0.149379	0.019701	0.123443

Similarity scores between patent #1 and others

Implementation : Content-based Method (Continued)

- **Results of content-based recommendation system testing**

- Recommending patents by keywords “data analytics” - a detailed list of patents that have highest similarity with keywords

```
tfidf_weighted_words('data analytics')
```

	id	date	abstract	title	kind	num_claims
3636	7352993	2008-04-01	A data reproducing system of the invention inc...	Data reproducing apparatus and data reproducin...	B2	20.0
3938	7656400	2010-02-02	A method and a device for converting data of a...	Image data editing device and method, and imag...	B2	24.0
3324	7039656	2006-05-02	A system for synchronizing data records betwee...	Method and apparatus for synchronizing data re...	B1	13.0
2951	6665283	2003-12-16	A wireless communication system transmits data...	Method and apparatus for transmitting data in ...	B2	17.0

Implementation : Content-based Method (Continued)

- **Results of content-based recommendation system testing**

- Recommending similar patents by contents based on a selected patent "*Image reproducing apparatus*" (ID: 6728471)

```
tfidf_similarity(6728471)
```

	id	date	abstract	title	kind	num_claims
982	4683500	1987-07-28	A method for reproducing picture images in wh...	Method for reproducing picture image	A	3.0
1384	5086358	1992-02-04	An apparatus for recording and reproducing in...	Recording and reproducing apparatus	A	11.0
3636	7352993	2008-04-01	A data reproducing system of the invention inc...	Data reproducing apparatus and data reproducin...	B2	20.0
4201	7920452	2011-04-05	Provided are a recording/reproducing apparatus...	Recording/reproducing apparatus and method	B2	12.0
1973	5680486	1997-10-21	An image processor includes an image memory w...	Image processing apparatus	A	16.0

Implementation : Knowledge-based Method

- Input features: user requirements (constraints), item features
- Output: mediating between user preferences and item properties and deriving a set of patents that fulfill the constraints

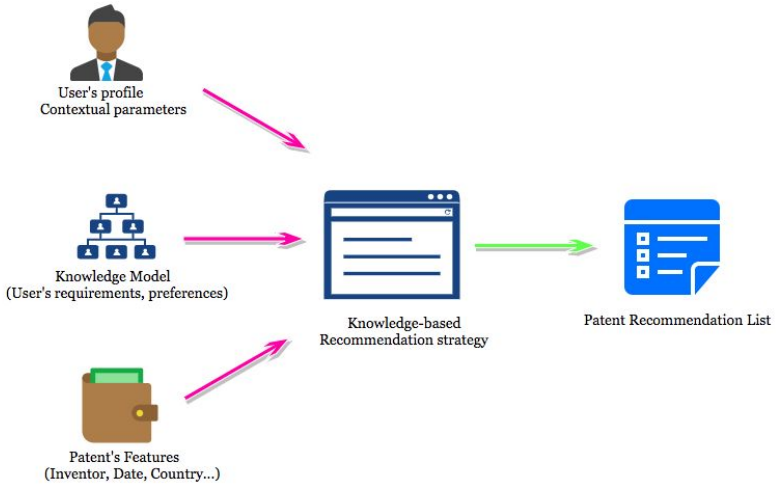
	id	type	country	date	abstract	title	A	B	C	D	E	F	G	H	Y
0	10000000	utility	US	2018-06-19	A frequency modulated (coherent) laser detecti...	Coherent LADAR using intra-pixel quadrature de...	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
1	10001015	utility	US	2018-06-19	Airfoil and hydrofoil systems include structur...	Drag reduction systems having fractal geometry...	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
2	10002022	utility	US	2018-06-19	A method, a computer program product, and a co...	Processing interrupt requests	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
3	10003026	utility	US	2018-06-19	A ladder tetrazine polymer is disclosed.	Ladder tetrazine polymers	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0
4	10004044	utility	US	2018-06-19	[Object] To achieve both prevention of harmful...	Communication control apparatus and wireless c...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Patent information: id, country, category...

	patent_id	inventor_id	name_first	name_last
16493759	T998013	T998012-3	Helen C. F.	Su
16493760	T998012	T998012-4	Roy E.	Bry
16493761	T998013	T998012-4	Roy E.	Bry
16493762	T998012	T998012-5	Robert	Davis
16493763	T998013	T998012-5	Robert	Davis

	patent_id	lawyer_id	name_first	name_last	organization
7848624	10186949	f096b46361f6143fe8d1cb5050409547	Daniel	Morris, Esq.	NaN
7848625	4371738	5cb9ef19159e15dddaee5117bc9b2e92	Birgie E.	Morris	NaN
7848626	8374858	d649c64db7754d849acad2c5c8c651a0	Guarav	Mohindra	NaN
7848627	7898229	55199546539b052e1389d0cca8601702	NaN	NaN	Thompson Patent Law Offices, PC
7848628	4296604	efc85be5833eaf435f6c1b91d2eb32a5	NaN	NaN	Finnegan, Henderson, Farabow, et al.

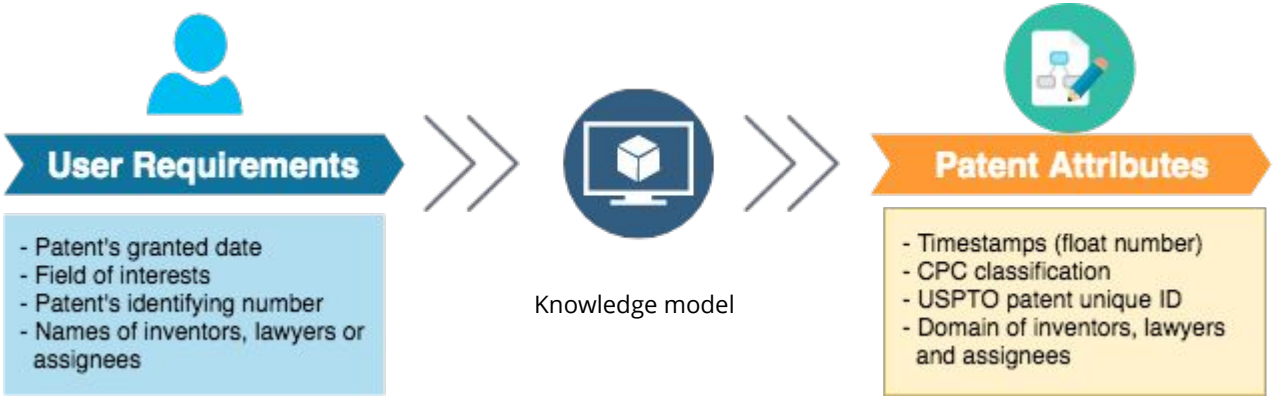
Information of inventors, lawyers, assignees



Implementation : Knowledge-based Method (Continued)

- Knowledge-based method - constraint based model
- Mechanism: CSP(constraint satisfaction problem) - knowledge-based recommendation system with declarative knowledge representation

$$CSP(X_I \cup X_U, D, SRS \cup KB \cup I)$$



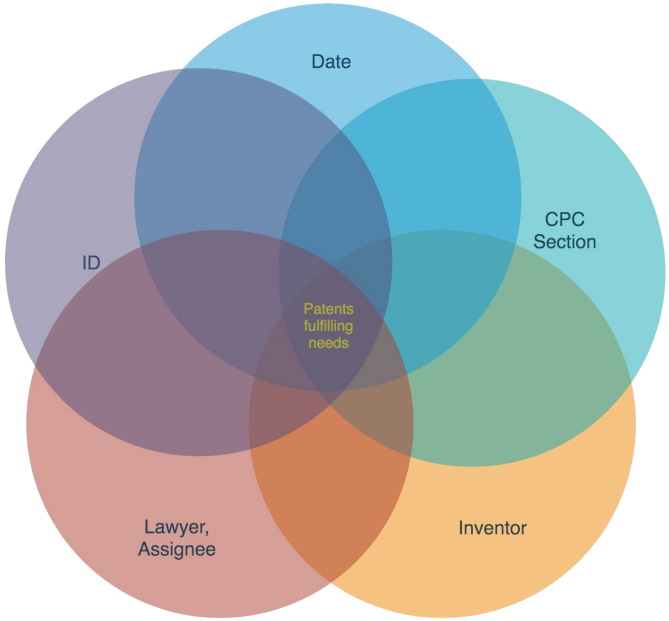
Implementation : Knowledge-based Method (Continued)

- **Results of constraint-based recommendation system testing**

Recommending patents based on user's requirements - "patents granted this year, in the area of Electricity."

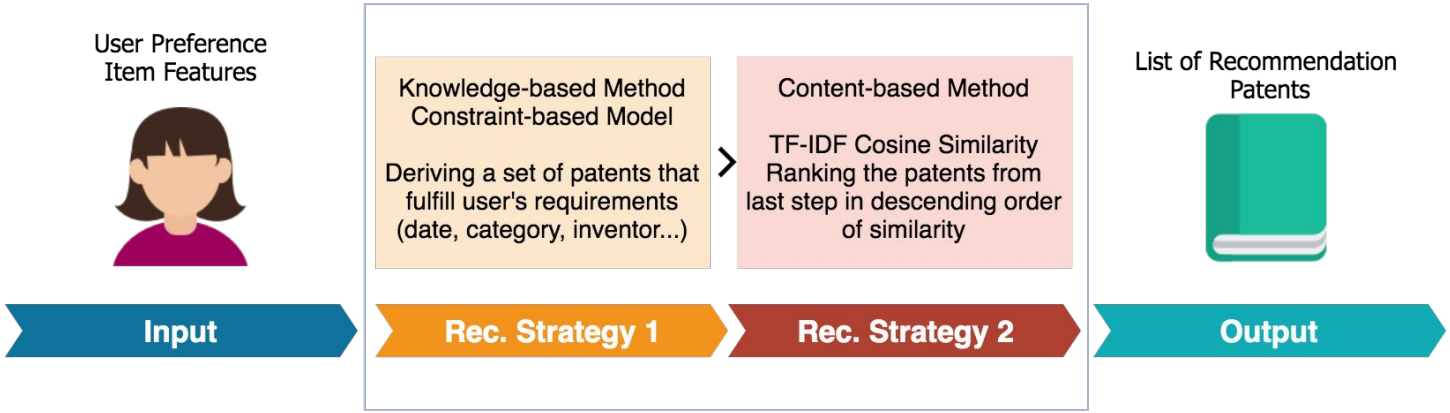
```
search_patent(None, '2019-01-01', None, None, 'H', None, None, None)
```

	id	date	abstract	title	kind	num_claims	A	B	C	D	E	F	G	H	Y
176	10177267	2019-01-08	An UV photodetector includes: a substrate, a t...	Photodetector	B2	14.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0
181	10182289	2019-01-15	An earpiece (100) and acoustic management modu...	Method and device for in ear canal echo suppre...	B2	12.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN
183	10184301	2019-01-22	An embodiment includes a downhole tool with fi...	Downhole drilling tools and connection system ...	B2	30.0	NaN	NaN	NaN	NaN	1.0	NaN	NaN	1.0	NaN
184	10185309	2019-01-22	In one embodiment, a tangible, non-transitory ...	Systems and methods for recommending component...	B2	23.0	NaN	NaN	NaN	NaN	NaN	1.0	1.0	1.0	1.0



Implementation : Hybrid - Pipelined

- Combination of content-based strategy and knowledge-based strategy
- First constraint-based system excludes patents that don’t fulfill the user’s requirements (category, date, inventor names…), and the second content-based recommender assigns scores for patents based on context features.



Hybrid System - Pipelined

Implementation : Hybrid - Pipeline (Continued)

- **Results of pipelined hybridization system testing**

Test example:

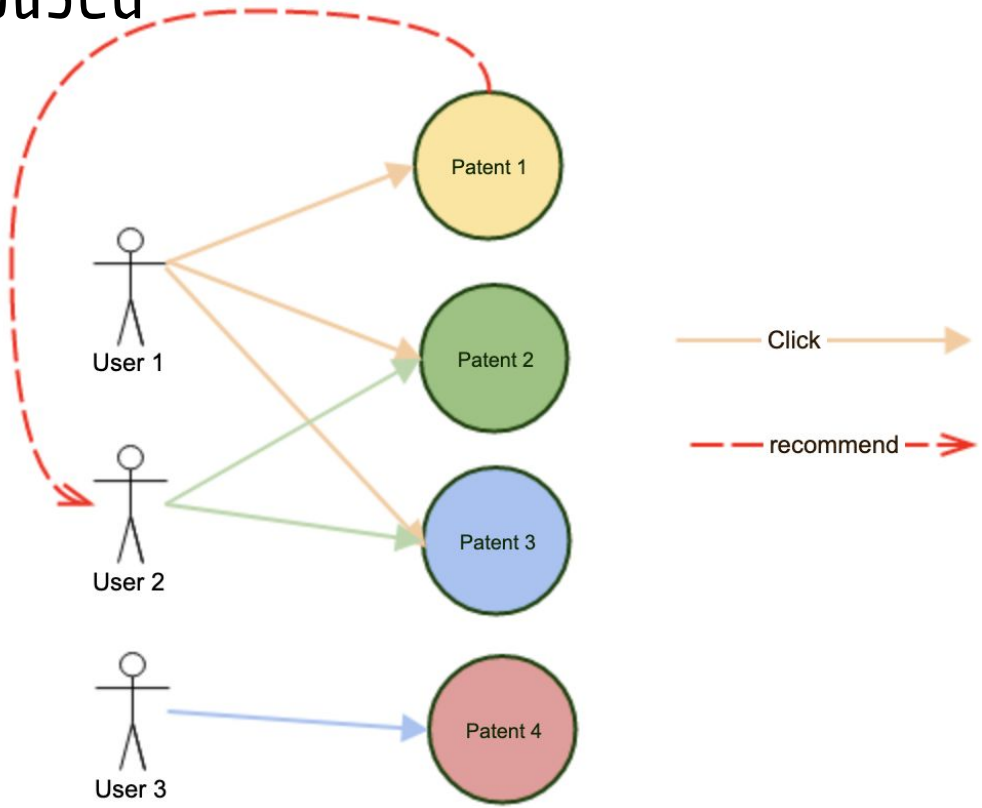
Recommending patents that satisfy the requirements: granted after 1995, and inventor(s) is called “Joseph”. The content of patents should be related to “laser”.

```
hybrid_pipe(None, '1995-01-01', None, None, None, 'Joseph', None, None, 'laser')
```

	id	date	abstract	title	kind	num_claims	A	B	C	D	E	F	G	H	Y	inventor_name	lawyer_name
0	10000000	2018-06-19	A frequency modulated (coherent) laser detecti...	Coherent LADAR using intra-pixel quadrature de...	B2	20.0	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	['Joseph C. Marron']	[' ,Munck Wilson Mandala, LLP']
2614	6327090	2001-12-04	A system which is preferably employed in a la...	Multiple laser beam generation	A	25.0	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0	NaN	['Joseph F. Rando', 'Timothy J. Litvin']	['Thomas M. Freiburger,']
2919	6633196	2003-10-14	An integrated circuit die includes a bond pad ...	Device and method for limiting the extent to w...	B2	31.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	['Joseph C. Sher']	[' ,TraskBritt']
3448	7164096	2007-01-16	A method for the fabrication of large metal ma...	Continuous metal matrix composite consolidation	B1	13.0	NaN	1.0	1.0	NaN	NaN	NaN	NaN	NaN	1.0	['Brian E. Joseph', 'Brian L. Gordon', 'James ...	['Philip D. Lane,']

Implementation : CF User-based

1. Store Click/View history data for each login user
2. Calculate pairwise Pearson correlation between users
3. Get the nearest neighbor
4. Recommend nearest neighbor's viewed patents to user (unviewed by this user)

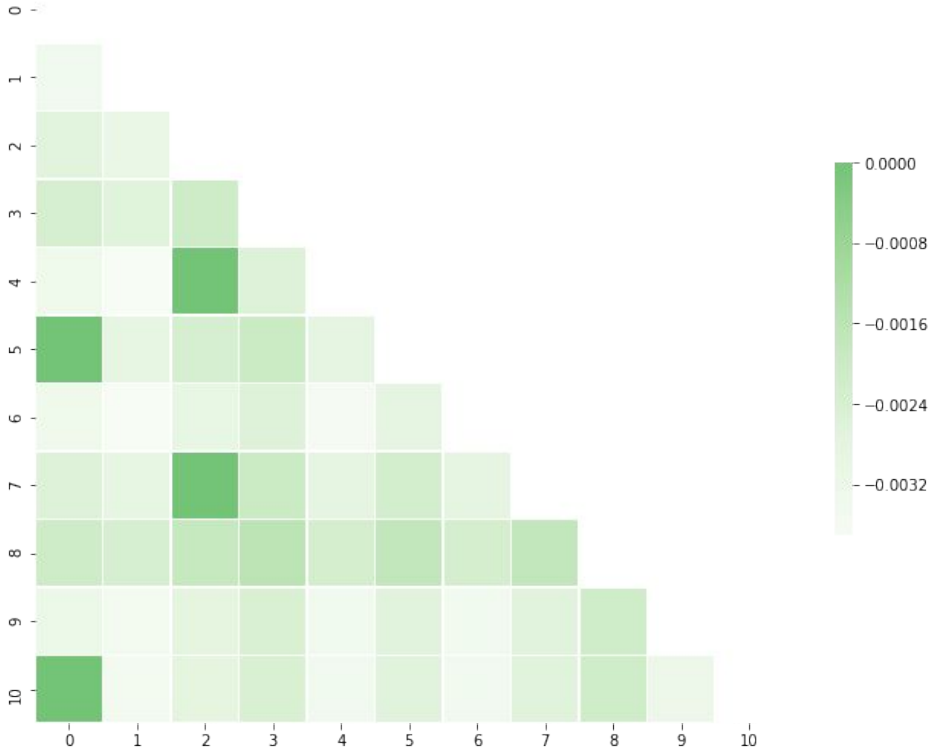


Implementation : CF User-based

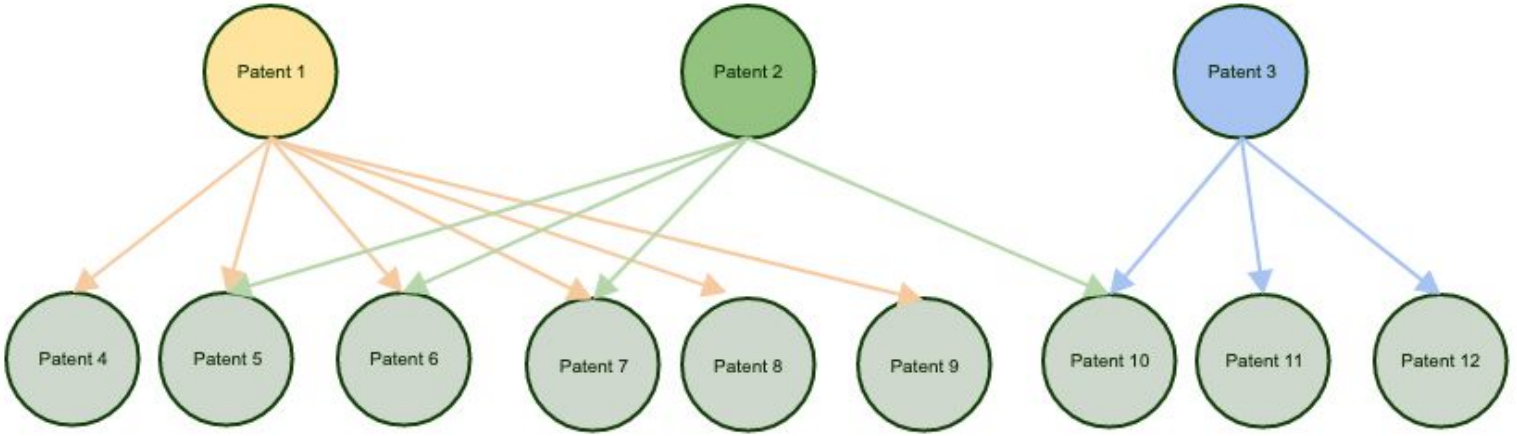
user1: {4032689, 4529347, 4797688...}

user2: {4919865, 4926875, 5423670...}

Pearson Correlation Matrix



Implementation : CF Item-based



Jaccard Similarity based on Citations

similarity(patent1, patent2) > similarity(patent2, patent3)

Implementation : CF Item-based

- One-Hot Encoding

	patent_1	patent_2	patent_3	patent_4	.....	patent_(n-2)	patent_(n-1)	patent_n
patent_1	0	1	0	0	.....	0	0	1
patent_2	0	0	0	1	0	0	0	0

- High-Dimensional Sparse Data : scipy.sparse.csr_matrix

```
citations of patent: 3943599
```

	patent_id	citation_id
62757	3943599	1452098
62758	3943599	2004581

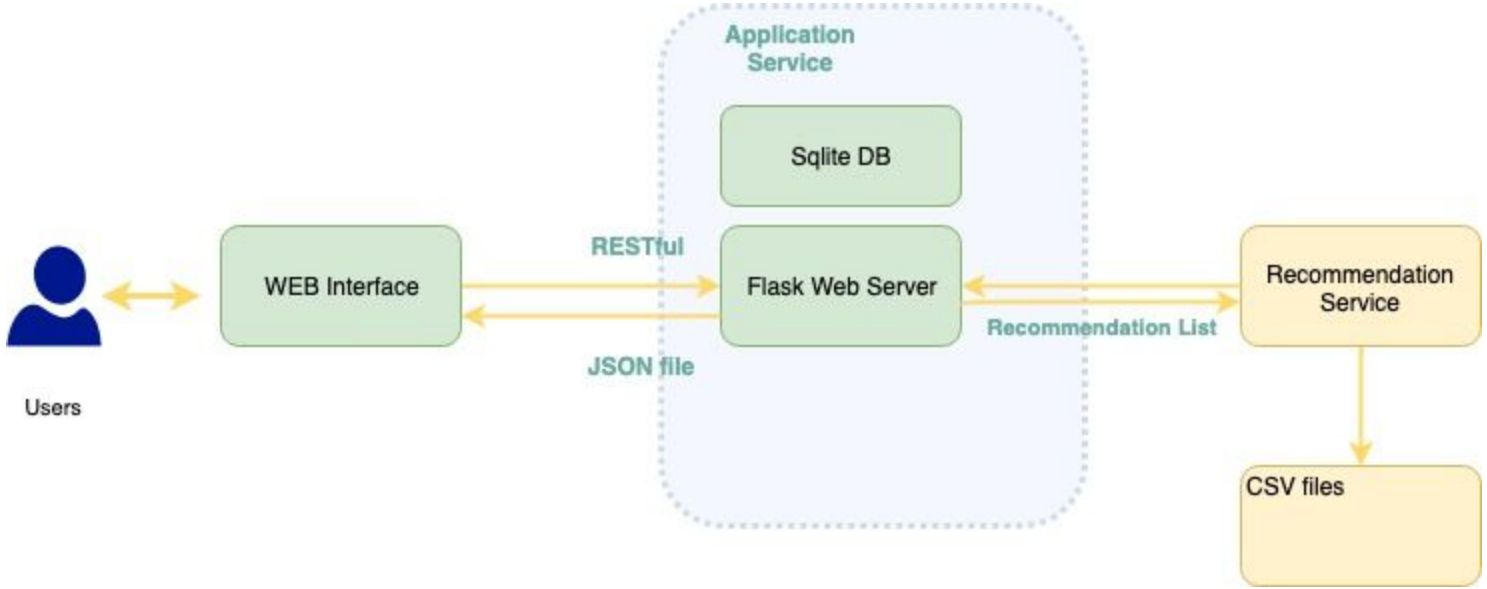
```
citations of patent: 4079741
```

	patent_id	citation_id
61069	4079741	2004581
61070	4079741	2900661
61071	4079741	1743590
61072	4079741	2083380
61073	4079741	1232617

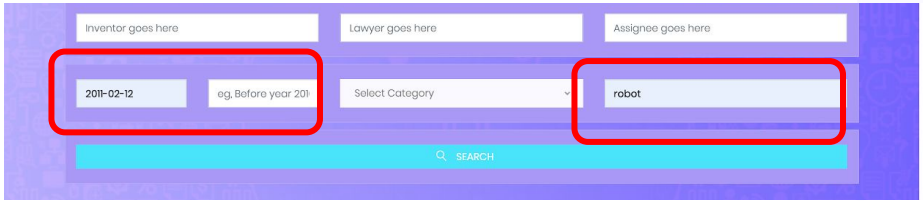
```
:
index1 = patent_index[patent_index.patent_id == patent_id_1].iloc[0]['patent_index']
index2 = patent_index[patent_index.patent_id == patent_id_2].iloc[0]['patent_index']
print("similarity between patent: %d and patent: %d is:" %(patent_id_1, patent_id_2))
print(sims[index1, index2])
```

```
similarity between patent: 3943599 and patent: 4079741 is:
0.16666666666666666
```

Application / User Interface



Hybrid Recommender - Pipeline Cascade



Top recommended patents based on hybrid recommendation

Method of manufacturing rotary scale, rotary scale, rotary encoder, driving apparatus, image pickup apparatus and robot apparatus

A method of manufacturing a rotary scale to be fixed to a rotating shaft of a rotating member includes a first step of forming, on a scale substrate, a scale pattern and a mark indicating an outer shape of the rotating shaft positioned such that a center axis of the rotating shaft coincides with a center axis of the scale pattern, a second step of cutting a first area of the scale substrate including the mark and having a first width, and a third step of cutting a second area including the mark that remains after the cutting of the first area, having a length in a circumferential direction of the scale substrate shorter than that in the first area and having a second width narrower than the first width.

Inventors: Masahiko Igaki', 'Haruhiko Horiguchi

Date: 2018-10-09

yers: Canon USA, Inc. I.P. Division

gnee: Canon Precision Inc.

ITENT BASED SIMILAR PATENTS | ITEM BASED SIMILAR PATENTS

Component integration apparatus and method for collaboration of heterogeneous robot

Provided is a technique that enables a robot to be remotely controlled (by a server) and enables a robot component to access an external component (a component of a server) in order for cooperation of heterogeneous robots operating on the basis of different component models. A component integration apparatus for collaboration of a heterogeneous robot according to an embodiment of the present invention comprises: a standard interface unit that provides a common standard interface for controlling components that control the individual functions of the robot; an adapter component that transmits commands to enable external components to call the components through the standard interface unit; and a proxy component that transmits commands to enable the components to call the external components through the standard interface unit.

Inventors: Hyun Kim', 'Kang-Woo Lee', 'Young-Ho Suh

Date: 2014-02-25

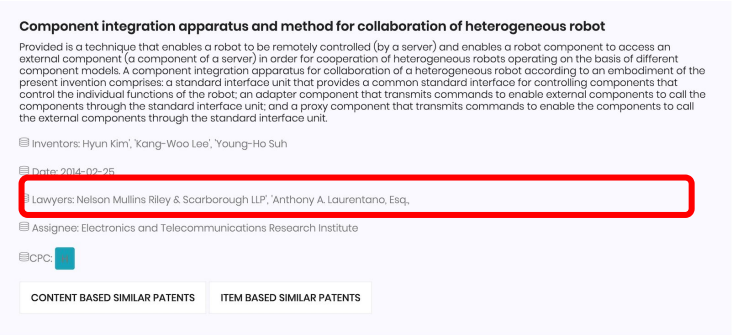
Lawyers: Nelson Mullins Riley & Scarborough LLP', 'Anthony A. Laurentano, Esq.,

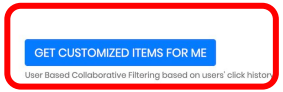
Assignee: Electronics and Telecommunications Research Institute

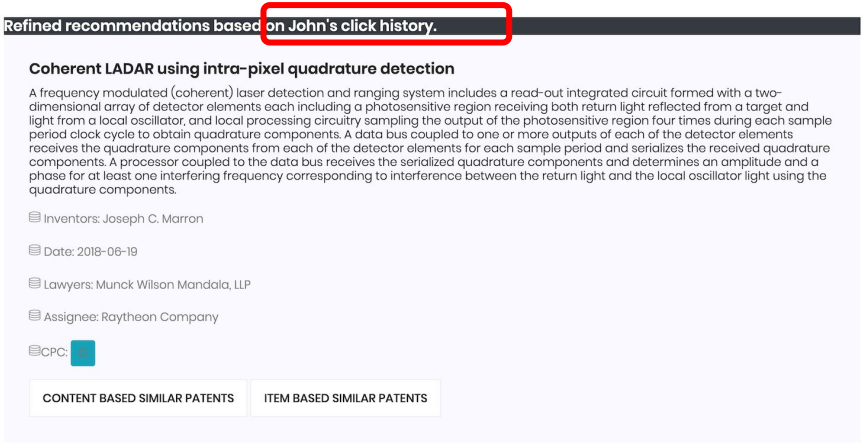
CPC:

CONTENT BASED SIMILAR PATENTS | ITEM BASED SIMILAR PATENTS

User Based CF Recommendation Based on Users' Clicking History







Step 1, Log in using username and password
Step 2, View detail of patent by clicking on patent title
Step 3, Click "GET CUSTOMIZED" button on the bottom of the page to store click history and get customized user-based recommendations

CONTENT-Based Recommender (TF-IDF)

Top recommended patents based on hybrid recommendation

Method of manufacturing rotary scale, rotary scale, rotary encoder, driving apparatus, image pickup apparatus and robot apparatus

A method of manufacturing a rotary scale to be fixed to a rotating shaft of a rotating member includes a first step of forming, on a scale substrate, a scale pattern and a mark indicating an outer shape of the rotating shaft positioned such that a center axis of the rotating shaft coincides with a center axis of the scale pattern, a second step of cutting a first area of the scale substrate including the mark and having a first width, and a third step of cutting a second area including the mark that remains after the cutting of the first area, having a length in a circumferential direction of the scale substrate shorter than that in the first area and having a second width narrower than the first width.

Inventors: Masahiko Igaki', 'Haruhiko Horiguchi

Date: 2018-10-09

Lawyers: Canon USA, Inc. I.P. Division

Assignee: Canon Precision Inc.

CPC:

CONTENT BASED SIMILAR PATENTS | ITEM BASED SIMILAR PATENTS

milar patents for patent id 1009488 based on Content(TF_IDF)

Rotary device for use in an engine

A rotary device for an engine includes a stator and a rotor concentric with and rotatable about an axis with respect to the stator. The rotor and the stator cooperate to provide a working chamber. A plurality of vanes are supported for radial movement on one of the stator and the rotor. Fluid is taken into the working chamber through an intake port and exhausted from the working chamber through an exhaust port. A biasing device biases each of the vanes to seal against one of the stator and the rotor. An actuator moves each of the vanes radially against the biasing device to a retracted position to vary a thermodynamic cycle of the rotary device as the rotor rotates with respect to the stator.

Inventors: Gilbert S. Staffend

Date: 2009-07-07

Lawyers: Dickinson Wright PLLC

Assignee: o

CPC:

CONTENT BASED SIMILAR PATENTS | ITEM BASED SIMILAR PATENTS

Rotary seal assembly

A seal assembly for a vehicle wheel assembly is provided for use between its two relatively rotatable, coaxial parts to seal the inner bearing chamber. The seal assembly includes a pair of nested casings forming an S-shaped recess in cross section defining inner and outer annular C-shaped concavities. Each casing has a peripheral seal extending radially from a peripheral edge and flexing against the inside sealing surface of the opposite casing to form a seal interface along the circular line of contact. Inner and outer annular floating wafer seals are disposed within the recess formed by the two casings and are pressed against the adjacent sealing surface within their respective concavities. The opposite edges of each independent wafer seal form additional seal interfaces making six total. Each wafer seal is self centering within its respective concavity.

Inventors: Frederick E. Lederman

Date: 1996-02-20

Lawyers: trick M. Griffin,

Assignee: General Motors Corporation

CPC:

CONTENT BASED SIMILAR PATENTS | ITEM BASED SIMILAR PATENTS

Image pickup apparatus

An image pickup apparatus includes a multifocal optical system having at least two different focal lengths; an image pickup device for converting an optical image formed by the multifocal optical system into an image signal; a first image processor for forming an original image defined by an object image at each focal length of the multifocal optical system with the image signal received from the image pickup device; and a second image processor for trimming the original image formed by the first image processor. An object image of an angle-of-view corresponding to an intermediate focal length between the two focal lengths of the multifocal optical system is complemented by a trimmed image formed by the second image processor.

Inventors: Ryota Ogawa', 'Masahiro Oono', 'Masakazu Saori

Date: 2009-05-05

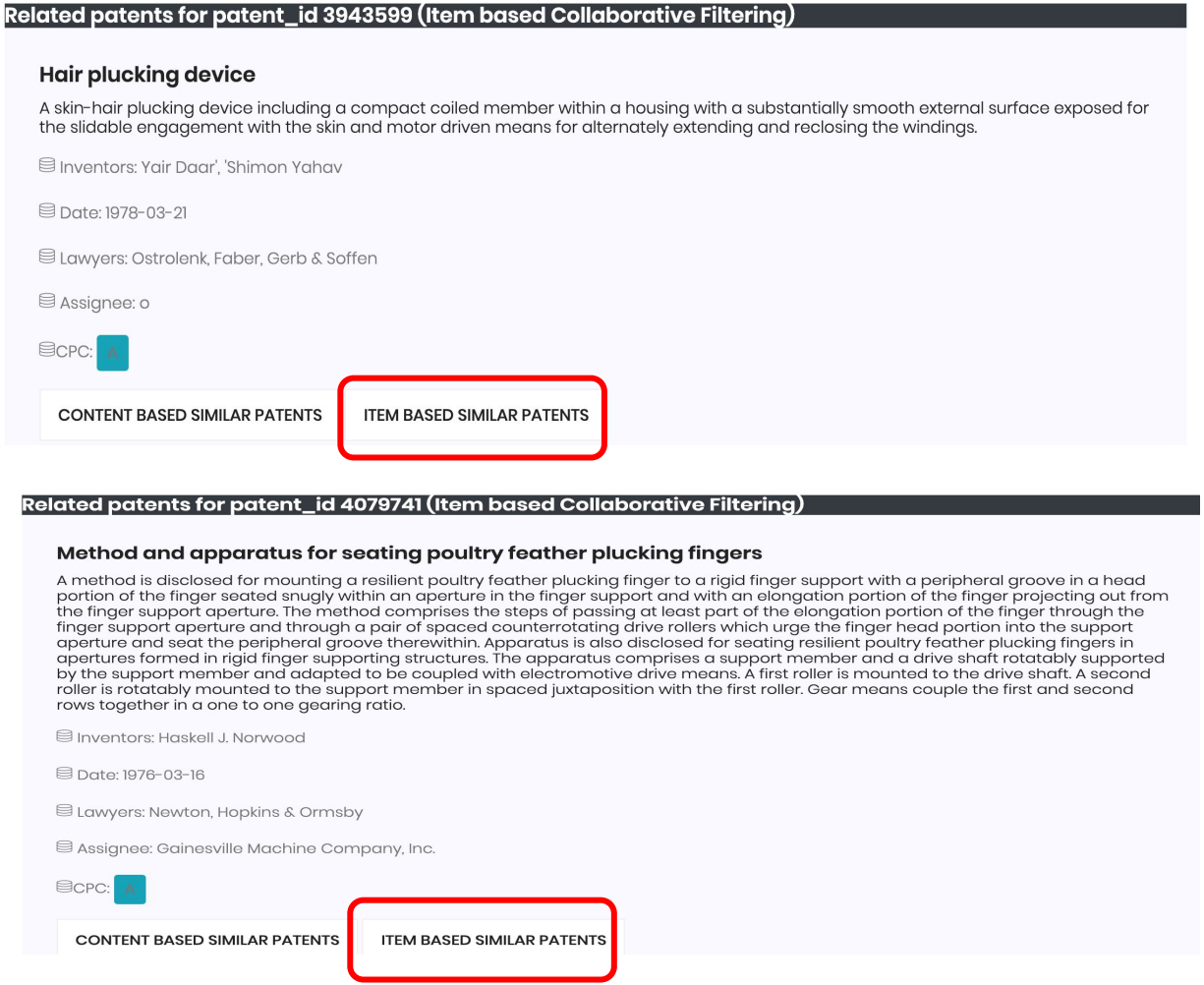
Lawyers: Greenblum & Bernstein, P.L.C.

Assignee: Hoya Corporation

CPC:

CONTENT BASED SIMILAR PATENTS | ITEM BASED SIMILAR PATENTS

Item-Based CF



Solution Evaluation

- Deliver a robust patent recommendation system using a wide variety of data analysis techniques (Hybrid pipelined system, Content-Based, Knowledge-Based, Collaborative Filtering, Cosine similarity, Pearson's Correlation, TF-IDF, etc)
- Able to recommend a set of patents that fit users' requirements as well as their personal taste
    - The user's profile in the system may have mixed results if some of their click data was for/from a friend or colleague

Impact after Implementation

- Searching results will be more precise and efficient
  - Obtaining the patents information that fulfill user requests and preferences
  - Automatically recommending patents based on user behaviours

- Reduce patent approval time

Thank You

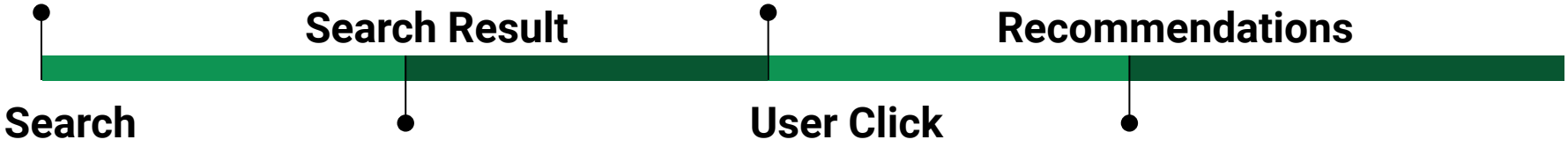
Appendix - System Design Chart

User Input Search Conditions

Date, Inventor Name, Assignee Name, Inventor Name, Keywords

User Click A Patent

User's click data will be stored and used for recommendations



Return Sorted Patent List

Keywords appear in title and abstract
Different Weight

Patent Recommendation Based on User's Click Events

Hybrid Recommenders

1. Content-based: TF-IDF, similarity between title / abstract
2. User-based Collaborative filtering
   Use user click data
3. Similarity from Citations: Jaccard Similarity