



Indian Institute of Technology, Guwahati

CS570 Assignment 3 Presentation

Prepared By:

Neelesh K Shukla
The Crawlers

Dated: 19th November 2016

CS 570

Quick Recap of Previous Assignments

Assignment 1 Gist

- Purpose of this assignment was to evaluate effectiveness of IR system.
- A small search engine was built using TFIDF and BM25 based retrieval.
- Precision@10 measure was used for evaluation.
- We concluded that mostly BM25 performs better than TF-IDF.

Assignment 1: The Crawlers' Search Engine

The Crawlers

☐ TFIDF ☐ BM25

Search Engine URL:

<http://172.16.115.152:8080/SearchEngine>

Indexer URL:

<http://172.16.115.152:8080/TheCrawlersIndexer>

Assignment 1 - Screenshots

TF-IDF Result

The Crawlers

☒ TFIDF ☐ BM25

78 results found.

[Category:Subnotebooks - Wikipedia, the free encyclopedia](#) score- 0.18399498
HP Mini HP Mini 2140 L **Lenovo** IdeaPad U300s **Lenovo** IdeaPad U310 M MSI Wind Netbook O OLPC XO One A110
[Zopo Mobile - Wikipedia, the free encyclopedia](#) score- 0.16263011
"Smartphone wars. Smartphone News: Q4 Results, **Lenovo** Grows, Vertu Goes Android, ZTE". BSN. BSN. Retrieved
[Laptotopical - Wikipedia, the free encyclopedia](#) score- 0.15934432
of brands including Apple Computer, Dell, **Lenovo** and Sony. The site has also written articles about
[Blade server - Wikipedia, the free encyclopedia](#) score- 0.14084181
to **Lenovo** in 2014.[22] Cisco offers a Unified Computing System in its blade server portfolio: a domain
[Kejian - Wikipedia, the free encyclopedia](#) score- 0.13942628
Konka Group Lemote **Lenovo** LeEco Meizu Panda Electronics Perception Digital Rockchip Smartisan
[Changhong - Wikipedia, the free encyclopedia](#) score- 0.13942628
Haier Hasee Hisense Hisense Kelon Huawei HiSilicon Ingenic Joyoung JXD Konka Group Lemote **Lenovo** LeEco
[Hisense Kelon - Wikipedia, the free encyclopedia](#) score- 0.13942628
Lenovo LeEco Meizu Panda Electronics Perception Digital Rockchip Smartisan Skyworth Supor SVA Group
[Haier - Wikipedia, the free encyclopedia](#) score- 0.13799623
Ingenic Joyoung JXD Konka Group Lemote **Lenovo** LeEco Meizu Panda Electronics Perception Digital
[TCL Corporation - Wikipedia, the free encyclopedia](#) score- 0.13799623
Group Lemote **Lenovo** LeEco Meizu Panda Electronics Perception Digital Rockchip Smartisan Skyworth Supor
[Oppo Electronics - Wikipedia, the free encyclopedia](#) score- 0.13799623
Hisense Kelon Huawei HiSilicon Ingenic Joyoung JXD Konka Group Lemote **Lenovo** LeEco Meizu Panda Electronics

BM25 Result

The Crawlers

☐ TFIDF ☒ BM25

78 results found.

[Zopo Mobile - Wikipedia, the free encyclopedia](#) score- 9.452357
"Smartphone wars. Smartphone News: Q4 Results, **Lenovo** Grows, Vertu Goes Android, ZTE". BSN. BSN. Retrieved
[Category:Subnotebooks - Wikipedia, the free encyclopedia](#) score- 9.325861
HP Mini HP Mini 2140 L **Lenovo** IdeaPad U300s **Lenovo** IdeaPad U310 M MSI Wind Netbook O OLPC XO One A110
[Haier - Wikipedia, the free encyclopedia](#) score- 9.17001
Ingenic Joyoung JXD Konka Group Lemote **Lenovo** LeEco Meizu Panda Electronics Perception Digital
[TCL Corporation - Wikipedia, the free encyclopedia](#) score- 9.17001
Group Lemote **Lenovo** LeEco Meizu Panda Electronics Perception Digital Rockchip Smartisan Skyworth Supor
[Oppo Electronics - Wikipedia, the free encyclopedia](#) score- 9.17001
Hisense Kelon Huawei HiSilicon Ingenic Joyoung JXD Konka Group Lemote **Lenovo** LeEco Meizu Panda Electronics
[Meizu - Wikipedia, the free encyclopedia](#) score- 9.17001
Ingenic Joyoung JXD Konka Group Lemote **Lenovo** LeEco Meizu Panda Electronics Perception Digital
[IBM - Wikipedia, the free encyclopedia](#) score- 9.146968
businesses to **Lenovo** (2005 and 2014, respectively), and acquiring companies such as PwC Consulting (2002
[Blade server - Wikipedia, the free encyclopedia](#) score- 9.084351
to **Lenovo** in 2014.[22] Cisco offers a Unified Computing System in its blade server portfolio: a domain
[Laptotopical - Wikipedia, the free encyclopedia](#) score- 8.92915
of brands including Apple Computer, Dell, **Lenovo** and Sony. The site has also written articles about
[Compaq - Wikipedia, the free encyclopedia](#) score- 8.9112215
PC division to **Lenovo**. The Inquirer noted that the continued low return on investment and small

TF-IDF P@10:

5/10

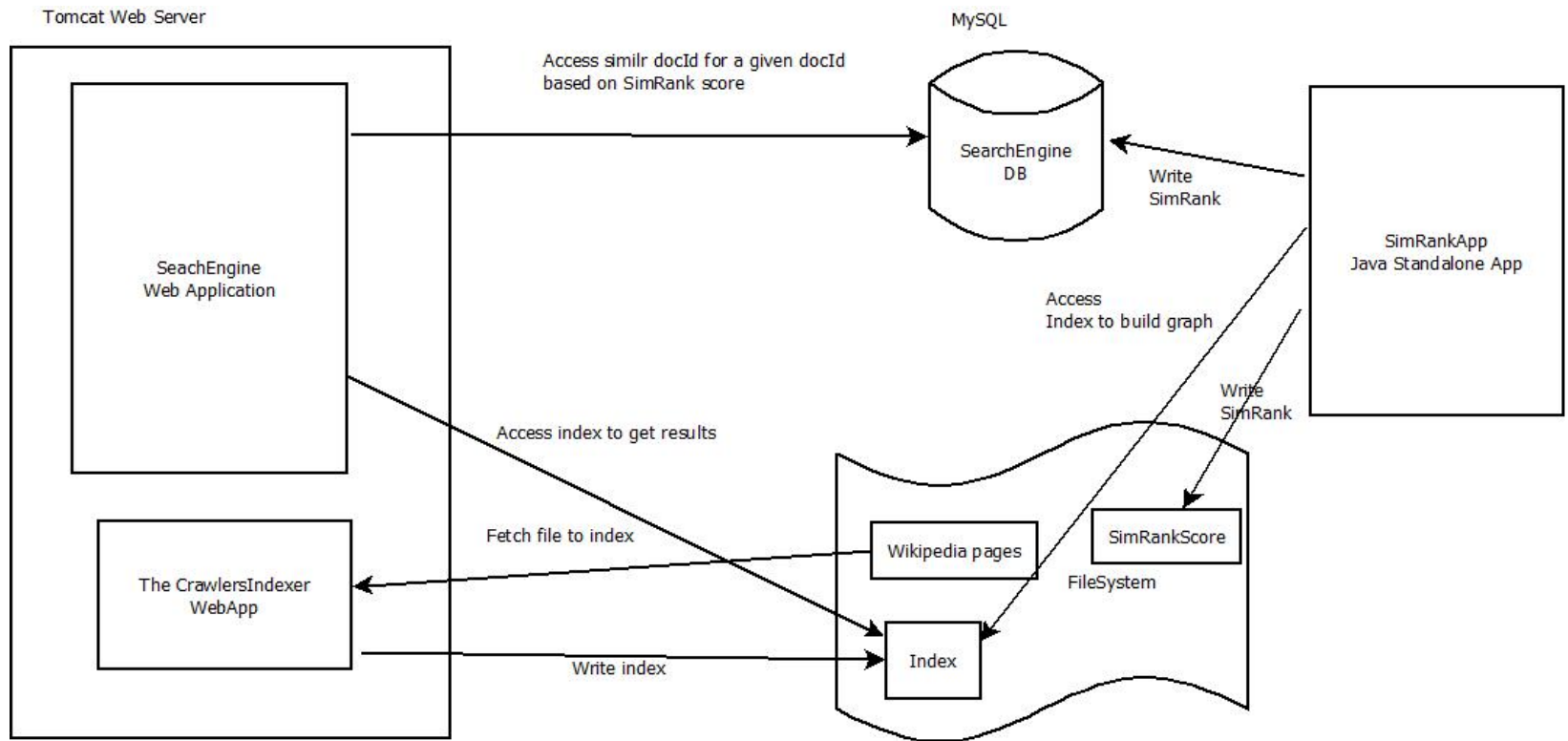
BM25 P@10:

6/10

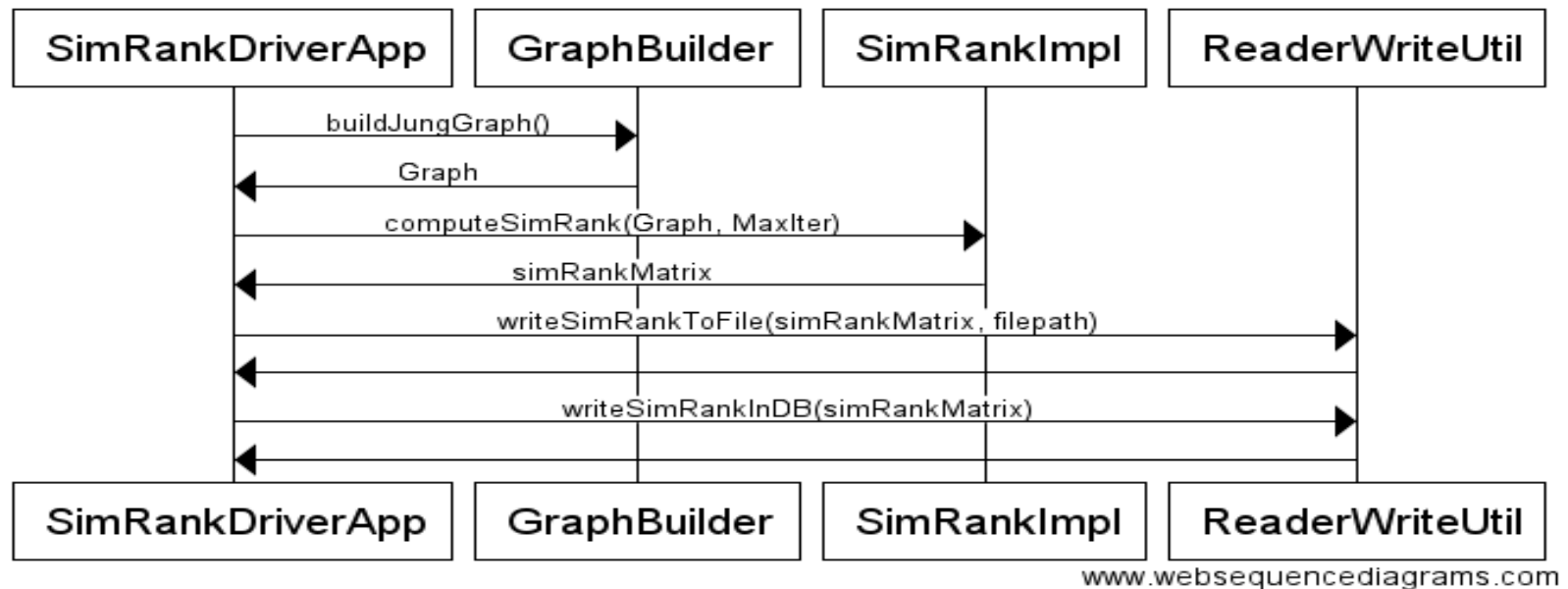
Assignment 2 Gist

- In this assignment we worked on structural contextual similarity measure called SimRank.
- Our application used SimRank to provide searches related to the top result returned by BM25.
- Cosine similarity was used to define relation between two documents, based on following assumptions.
 - Two documents are similar if they are referred by document which has similar content.
 - If two documents have similar content, they are somehow related to similar topics/areas.
- We built assignment 2 on the top of assignment 1.

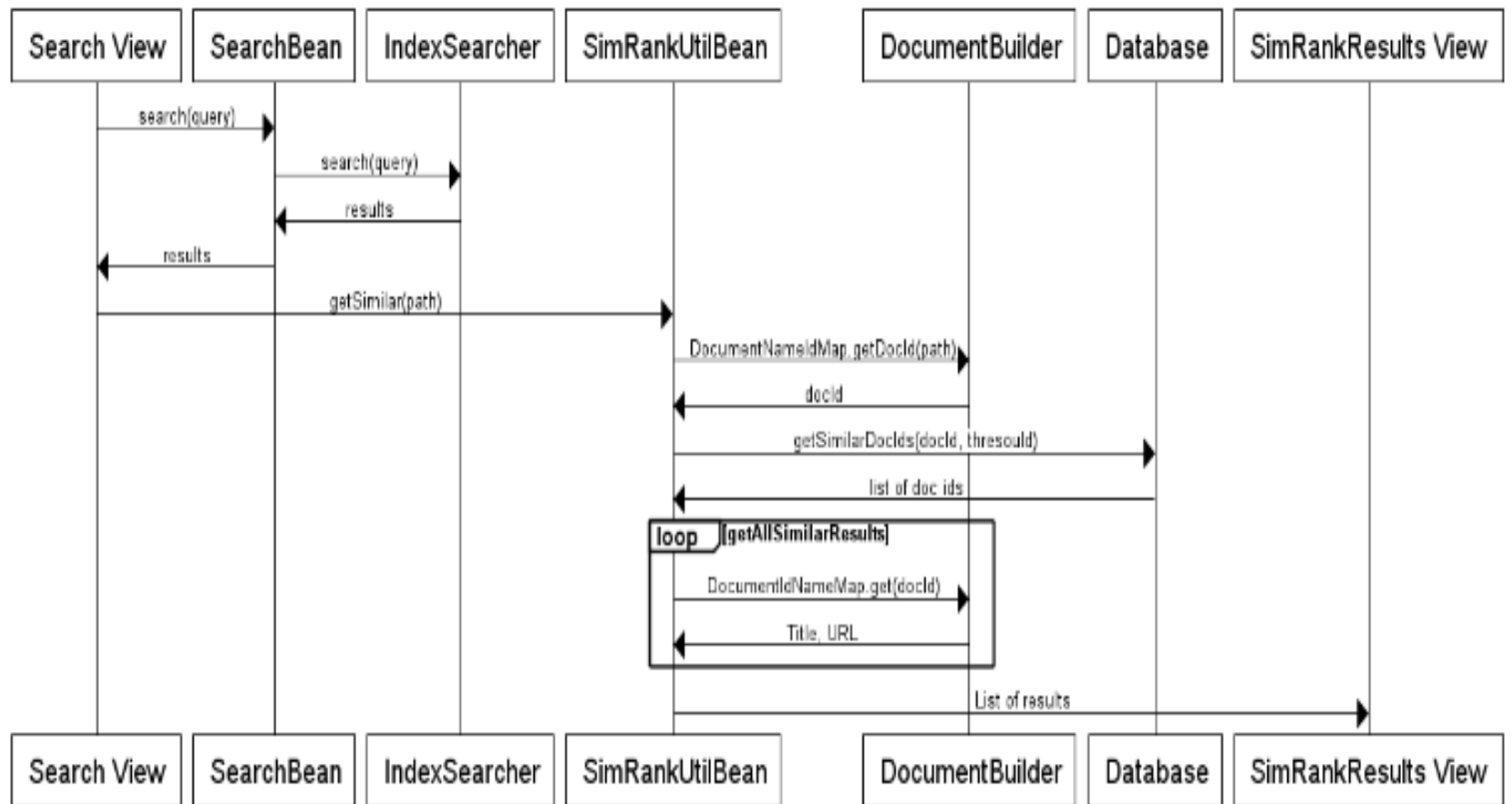
Assignment 2: System Overview



Assignment 2: SimRankApp Sequence Diagram



Assignment 2: Search Engine Sequence Diagram



Assignment 2: Screenshots – Search Step 1

The Crawlers

518 results found.

[Get the similar results to highest ranked result using SimRank](#)

[Liquid Propulsion Systems Centre - Wikipedia, the free encyclopedia](#)

=710032476" Categories: Space programme of **India** Space technology **research** institutes Rocket engine

[Shobhana Narasimhan - Wikipedia, the free encyclopedia](#)

Scientific **Research** in Bangalore, **India**. Her main area of interest is computational nanoscience. Her

[Gateway of India - Wikipedia, the free encyclopedia](#)

Bank of **India** Mint Dalal Street Nariman Point Education and **Research** Universities & Colleges

[Syed I. Ahson - Wikipedia, the free encyclopedia](#)

in the United Kingdom. He was a National Merit Scholarship Award, **India** (1959–1965). He also won

[Rajiv Gandhi International Cricket Stadium - Wikipedia, the free encyclopedia](#)

Stadium Ground information Location Greater Hyderabad, Telangana, **India** Establishment 2003 Capacity

[Nehru Stadium, Pune - Wikipedia, the free encyclopedia](#)

: **India** v England Last ODI 3 November 2005: **India** v Sri Lanka Team information Maharashtra (1969

[PSLV-C34 - Wikipedia, the free encyclopedia](#)

". Indian Space **Research** Organisation. Retrieved 22 June 2016. ^ a b c d "**India** sets new record in space

[Indian Institute of Technology Kharagpur - Wikipedia, the free encyclopedia](#)

of SIDBI (Small Industries Development Bank of **India**).[40] Sponsored **research**[edit] Kalpana Chawla

[PSLV-D1 - Wikipedia, the free encyclopedia](#)

and the mission was a total failure.[1][4] See also[edit] Indian Space **Research** Organisation Polar

[Hartebeesthoek Radio Astronomy Observatory - Wikipedia, the free encyclopedia](#)

Hartebeesthoek Radio Astronomy Observatory DSS 51 at Hartebeesthoek Organization National **Research** Foundation

[Lakshmibai National College of Physical Education - Wikipedia, the free encyclopedia](#)

, **India**, which bears the name of Lakshmibai, Rani of Jhansi, a heroine of the First War of Independence

[Satish Dhawan Space Centre SLV Launch Pad - Wikipedia, the free encyclopedia](#)

in Sriharikota, **India** that began operation in 1979 and was decommissioned in 1994. It was used

Assignment 2: Screenshots - Search Step 2

The Crawlers

50 similar results found using SIMRANK.

[Liquid Propulsion Systems Centre - Wikipedia, the free encyclopedia](#) score-1.0
[PSLV-D1 - Wikipedia, the free encyclopedia](#) score-0.215945
[Satish Dhawan Space Centre SLV Launch Pad - Wikipedia, the free encyclopedia](#) score-0.173943
[PSLV-C34 - Wikipedia, the free encyclopedia](#) score-0.0830573
[Geosynchronous Satellite Launch Vehicle Mark III - Wikipedia, the free encyclopedia](#) score-0.0728792
[Geosynchronous Satellite Launch Vehicle Mark III - Wikipedia, the free encyclopedia](#) score-0.072869
[ANUSAT - Wikipedia, the free encyclopedia](#) score-0.0611716
[Saturn I SA-1 - Wikipedia, the free encyclopedia](#) score-0.0406297
[IRS-P1 - Wikipedia, the free encyclopedia](#) score-0.0401041
[Long March \(rocket family\) - Wikipedia, the free encyclopedia](#) score-0.0372916
[INSAT-2DT - Wikipedia, the free encyclopedia](#) score-0.0320474
[Cygnus CRS Orb-3 - Wikipedia, the free encyclopedia](#) score-0.0295831
[GSAT-16 - Wikipedia, the free encyclopedia](#) score-0.0271929
[IRNSS-1B - Wikipedia, the free encyclopedia](#) score-0.0209418
[IRNSS-1B - Wikipedia, the free encyclopedia](#) score-0.0209381
[Brahmagupta–Fibonacci identity - Wikipedia, the free encyclopedia](#) score-0.0205105
[Zenit-3SLB - Wikipedia, the free encyclopedia](#) score-0.020019
[GSAT-16 - Wikipedia, the free encyclopedia](#) score-0.018517
[Minotaur-C - Wikipedia, the free encyclopedia](#) score-0.0170561
[IRNSS-1G - Wikipedia, the free encyclopedia](#) score-0.0165307
[GSAT-17 - Wikipedia, the free encyclopedia](#) score-0.0159467
[GSAT-1 - Wikipedia, the free encyclopedia](#) score-0.0155644
[Long March 3 - Wikipedia, the free encyclopedia](#) score-0.014947
[IRS-1B - Wikipedia, the free encyclopedia](#) score-0.0148951
[IRNSS-1F - Wikipedia, the free encyclopedia](#) score-0.0141603
[IRNSS-1F - Wikipedia, the free encyclopedia](#) score-0.0141322
[GSAT-10 - Wikipedia, the free encyclopedia](#) score-0.0139188
[GSAT-19E - Wikipedia, the free encyclopedia](#) score-0.0133025
[GSAT-19E - Wikipedia, the free encyclopedia](#) score-0.0132817
[Titan 34D - Wikipedia, the free encyclopedia](#) score-0.0131057

Assignment - 3

Assignment 3 : Directions ->

- ✗ Use some API to improve search engine. (Application Development)
- ✗ Take existing ideas and implement them to improve search experience.
- ✓ Come up with own idea based on our previous observations and learning. Be it a small improvement.

Assignment 3 : Improvements

- Devised a new ranking function combining SimRank and BM25.

The Crawlers

☐ TFIDF ☐ BM25 

The Crawlers

☐ TFIDF ☐ BM25 ☐ SIMBM25

Assignment 3 : Improvements

- Provided 'Similar' results feature for each result.

The Crawlers

518 results found.

[Get the similar results to highest ranked result using SimRank](#)

[Liquid Propulsion Systems Centre - Wikipedia, the free encyclopedia](#)
=710032476" Categories: Space programme of **India** Space technology **research** institutes Rocket engine
[Shobhana Narasimhan - Wikipedia, the free encyclopedia](#)
Scientific **Research** in Bangalore, **India**. Her main area of interest is computational nanoscience. Her
[Gateway of India - Wikipedia, the free encyclopedia](#)
Bank of **India** Mint Dalal Street Nariman Point Education and **Research** Universities & Colleges
[Syed I. Ahson - Wikipedia, the free encyclopedia](#)
in the United Kingdom. He was a National Merit Scholarship Award, **India** (1959–1965). He also won
[Rajiv Gandhi International Cricket Stadium - Wikipedia, the free encyclopedia](#)
Stadium Ground information Location Greater Hyderabad, Telangana, **India** Establishment 2003 Capacity
[Nehru Stadium, Pune - Wikipedia, the free encyclopedia](#)
: **India** v England Last ODI 3 November 2005: **India** v Sri Lanka Team information Maharashtra (1969
[PSLV-C34 - Wikipedia, the free encyclopedia](#)
". Indian Space **Research** Organisation. Retrieved 22 June 2016. ^a ^b ^c ^d "**India** sets new record in space
[Indian Institute of Technology Kharagpur - Wikipedia, the free encyclopedia](#)
of SIDBI (Small Industries Development Bank of **India**).^[40] Sponsored **research**^[edit] Kalpana Chawla
[PSLV-D1 - Wikipedia, the free encyclopedia](#)
and the mission was a total failure.^{[1][4]} See also^[edit] Indian Space **Research** Organisation Polar
[Hartebeesthoek Radio Astronomy Observatory - Wikipedia, the free encyclopedia](#)
Hartebeesthoek Radio Astronomy Observatory DSS 51 at Hartebeesthoek Organization National **Research** Foundation
[Lakshmbai National College of Physical Education - Wikipedia, the free encyclopedia](#)
, **India**, which bears the name of Lakshmbai, Rani of Jhansi, a heroine of the First War of Independence
[Satish Dhawan Space Centre SLV Launch Pad - Wikipedia, the free encyclopedia](#)
in Sriharikota, **India** that began operation in 1979 and was decommissioned in 1994. It was used



The Crawlers

☐ TFIDF ☐ BM25 ☐ SIMBM25

17 results found.

[Charles Daniels \(tenor\) - Wikipedia, the free encyclopedia](#)
Daniels is an English tenor, particularly noted for his performances of baroque **music**. He is a frequent
[Radhames Aracena - Wikipedia, the free encyclopedia](#)
Nationality Dominican Occupation radio host, **music** producer, businessman Radhames Aracena (13 May 1930
[Moulin Rouge \(band\) - Wikipedia, the free encyclopedia](#)
) is a Slovenian popular **music** group consisting of vocalist Alenka Šmid-Čena and songwriter, keyboardist
[KITS - Wikipedia, the free encyclopedia](#)
radio consultant Mike Joseph's Hot Hits Top-40 **music** format. Seven air personalities were recruited
[Long March \(disambiguation\) - Wikipedia, the free encyclopedia](#)
Literature 3.3 **Music** 4 See also Military actions and prisoner of war marches^[edit] Long March
[Urban - Wikipedia, the free encyclopedia](#)
contemporary, an African American **music** radio format Mainstream urban Urban adult contemporary Urban
[Andrew Belle - Wikipedia, the free encyclopedia](#)
-songwriter, [...] I wanted to make **music** like they were making." [3] Black Bear was released
[Peelander-Z - Wikipedia, the free encyclopedia](#)
) Space Vacation (2012) Metalander-Z (2013) Compilations^[edit] You Call This **Music**?! Volume 2 (2002
[Atonement Academy - Wikipedia, the free encyclopedia](#)
, social studies, foreign languages, English, religion, **music**, and physical education instruction
[Grand Ole Opry \(film\) - Wikipedia, the free encyclopedia](#)
Leon Weaver Frank Weaver June Weaver Lois Ranson Allan Lane Henry Kolker **Music** by William Lava
[Miracle of the White Suit - Wikipedia, the free encyclopedia](#)
Miguel Ángel Rodríguez Julia Martínez **Music** by Jesús Guridi Cinematography Cecilio Paniagua Edited
[Whiplash \(band\) - Wikipedia, the free encyclopedia](#)
DeLong, a Pennsylvania-Dutch native, with **music** influences from the 80's NY Metal/Hardcore/Punk scene
[John M. Grunsfeld - Wikipedia, the free encyclopedia](#)
, sailing, bicycling, and **music**. His father, Ernest Grunsfeld III, died in 2011 at the age of 81.^[5]
[Macedonian denar - Wikipedia, the free encyclopedia](#)
Architecture Cinema National emblem Cuisine Flag Historical coat of arms Literature **Music** Public
[Tune Book \(denar\) - Wikipedia, the free encyclopedia](#)

Assignment 3 : Improvements

- Better UI for listing SimRank based Results.
- 2-Step search to 1-step search.

The Crawlers

50 similar results found using SIMRANK.

[Liquid Propulsion Systems Centre - Wikipedia, the free encyclopedia](#) score-1.0
[PSLV-D1 - Wikipedia, the free encyclopedia](#) score-0.215945
[Satish Dhawan Space Centre SLV Launch Pad - Wikipedia, the free encyclopedia](#) score-0.173943
[PSLV-C34 - Wikipedia, the free encyclopedia](#) score-0.0830573
[Geosynchronous Satellite Launch Vehicle Mark III - Wikipedia, the free encyclopedia](#) score-0.0728792
[Geosynchronous Satellite Launch Vehicle Mark III - Wikipedia, the free encyclopedia](#) score-0.072869
[ANUSAT - Wikipedia, the free encyclopedia](#) score-0.0611716
[Saturn I SA-1 - Wikipedia, the free encyclopedia](#) score-0.0406297
[IRS-P1 - Wikipedia, the free encyclopedia](#) score-0.0401041
[Long March \(rocket family\) - Wikipedia, the free encyclopedia](#) score-0.0372916
[INSAT-2DT - Wikipedia, the free encyclopedia](#) score-0.0320474
[Cygnus CRS Orb-3 - Wikipedia, the free encyclopedia](#) score-0.0295831
[GSAT-16 - Wikipedia, the free encyclopedia](#) score-0.0271929
[IRNSS-1B - Wikipedia, the free encyclopedia](#) score-0.0209418
[IRNSS-1B - Wikipedia, the free encyclopedia](#) score-0.0209381
[Brahmagupta–Fibonacci identity - Wikipedia, the free encyclopedia](#) score-0.0205105
[Zenit-3SLB - Wikipedia, the free encyclopedia](#) score-0.020019
[GSAT-16 - Wikipedia, the free encyclopedia](#) score-0.018517
[Minotaur-C - Wikipedia, the free encyclopedia](#) score-0.0170561
[IRNSS-1G - Wikipedia, the free encyclopedia](#) score-0.0165307
[GSAT-17 - Wikipedia, the free encyclopedia](#) score-0.0159467
[GSAT-1 - Wikipedia, the free encyclopedia](#) score-0.0155644
[Long March 3 - Wikipedia, the free encyclopedia](#) score-0.014947
[IRS-1B - Wikipedia, the free encyclopedia](#) score-0.0148951
[IRNSS-1F - Wikipedia, the free encyclopedia](#) score-0.0141603
[IRNSS-1F - Wikipedia, the free encyclopedia](#) score-0.0141322
[GSAT-10 - Wikipedia, the free encyclopedia](#) score-0.0139188
[GSAT-19E - Wikipedia, the free encyclopedia](#) score-0.0133025
[GSAT-19E - Wikipedia, the free encyclopedia](#) score-0.0132817
[Titan 34D - Wikipedia, the free encyclopedia](#) score-0.0131057



The Crawlers

● TFIDF ● BM25 ● SIMBM25

17 results found.

[Charles Daniels \(tenor\) - Wikipedia, the free encyclopedia](#)

Daniels is an English tenor, particularly noted for his performances of baroque **music**. He is a frequent

[Radhames Aracena - Wikipedia, the free encyclopedia](#)

Nationality Dominican Occupation radio host, **music** producer, businessman Radhames Aracena (13 May 1930

[Moulin Rouge \(band\) - Wikipedia, the free encyclopedia](#)

) is a Slovenian popular **music** group consisting of vocalist Alenka Šmid-Čena and songwriter, keyboardist

[KITS - Wikipedia, the free encyclopedia](#)

radio consultant Mike Joseph's Hot Hits Top-40 **music** format. Seven air personalities were recruited

[Long March \(disambiguation\) - Wikipedia, the free encyclopedia](#)

Literature 3.3 **Music** 4 See also Military actions and prisoner of war marches[edit] Long March

[Urban - Wikipedia, the free encyclopedia](#)

contemporary, an African American **music** radio format Mainstream urban Urban adult contemporary Urban

[Andrew Belle - Wikipedia, the free encyclopedia](#)

-songwriter, [...] I wanted to make **music** like they were making." [3] Black Bear was released

[Peelander-Z - Wikipedia, the free encyclopedia](#)

) Space Vacation (2012) Metalander-Z (2013) Compilations[edit] You Call This **Music**?! Volume 2 (2002

[Atonement Academy - Wikipedia, the free encyclopedia](#)

, social studies, foreign languages, English, religion, **music**, and physical education instruction

[Grand Ole Opry \(film\) - Wikipedia, the free encyclopedia](#)

Leon Weaver Frank Weaver June Weaver Lois Ranson Allan Lane Henry Kolker **Music** by William Lava

[Miracle of the White Suit - Wikipedia, the free encyclopedia](#)

Miguel Ángel Rodríguez Julia Martínez **Music** by Jesús Guridi Cinematography Cecilio Paniagua Edited

[Whiplash \(band\) - Wikipedia, the free encyclopedia](#)

DeLong, a Pennsylvania-Dutch native, with **music** influences from the 80's NY Metal/Hardcore/Punk scene

[John M. Grunsfeld - Wikipedia, the free encyclopedia](#)

, sailing, bicycling, and **music**. His father, Ernest Grunsfeld III, died in 2011 at the age of 81. [5

[Macedonian denar - Wikipedia, the free encyclopedia](#)

Architecture Cúena National emblem Cuisine Flag Historical coat of arms Literature **Music** Public

[Tara Production - Wikipedia, the free encyclopedia](#)

Proposed Ranking Function: SIMBM25

- Structural contextual similarity plays role in understanding context.
- In assignment 2, we fetched results using BM25 and taken the 1st results as a reference to get the query intent and listed pages similar to topmost result.

BM25 and SimRank score can be combined to improve search result.

- We fetched new results which are similar to topmost BM25 result. If topmost result is biased then new search will be biased.

Use top K results of BM25 to calculate score.

$$\text{Score}(d) = \sum_{1 \leq i \leq k} \text{SimRankScore}(d_i, d)$$

Proposed Ranking Function: SIMBM25

➤ Previous scoring function has a flaw.

➤ Example

<code>SimRankScore(Res1, d1) = .10</code>	\rightarrow	<code>Score(d1) = .11</code>
<code>SimRankScore(Res8, d1) = .01</code>		
<code>SimRankScore(Res1, d2) = .01</code>	\rightarrow	<code>Score(d2) = .11</code>
<code>SimRankScore(Res8, d2) = .10</code>		

➤ Both d1 and d2 got same score but d1 must be ranked higher than d2.

Need to assign some weight to SimRankScore.

$$\text{Score}(d) = \sum w(d_i) * \text{SimRankScore}(d_i, d)$$

Proposed Ranking Function: SIMBM25

- BM25 score of the documents was used as weight.
- Final scoring function

$$\text{Score}(d) = \sum_{1 \leq i \leq k} \text{BM25Score}(d_i) * \text{SimRankScore}(d_i, d)$$

d_i = Document present in BM25 result at rank i

d = Document present in corpus to be ranked by SIMBM25 ranking function

*This is online scoring mechanism, scores are calculated on the fly.

Precision@10 Evaluation

Query: Research

BM25 Result

The Crawlers

☐ TFIDF ☒ BM25 ☐ SIMBM25

287 results found.

[Liquid Propulsion Systems Centre - Wikipedia, the free encyclopedia](#)

The Liquid Propulsion Systems Centre (LPSC; Hindi: द्रव नोदन प्रणाली केंद्र) is a **research**

[Sat Bir Singh Khalsa - Wikipedia, the free encyclopedia](#)

Massachusetts. Since 2007, Sat Bir Singh Khalsa has served as the Director of **Research** at both

[B. P. Loughridge - Wikipedia, the free encyclopedia](#)

at the Oklahoma Medical **Research** Foundation, and working as an exterminator for a pest control company. He

[PSLV-C34 - Wikipedia, the free encyclopedia](#)

a new record of deploying the maximum number of satellites by Indian Space **Research** Organisation

[Privadarshini Engineering College - Wikipedia, the free encyclopedia](#)

ensuring a high quality teaching learning and **research** environment. Motivated faculty - student

[Astronomy & Geophysics - Wikipedia, the free encyclopedia](#)

research, science policy, news, opinions, correspondence, and book and software reviews. Royal

[TVGuide.co.uk - Wikipedia, the free encyclopedia](#)

Media via Highbeam **Research**): 19 April 2007. (subscription required) ^ a b "Broadcasters seek legal

[National Sporting Goods Association - Wikipedia, the free encyclopedia](#)

Association Founded 1929 Type Trade Association Focus Education, **Research** & Advocacy Location Mount

[Mohammed Rafique Mughal - Wikipedia, the free encyclopedia](#)

of ethnoarchaeological **research** in Chitral, northern Pakistan. He has been responsible for the direction

[Satish Dhawan Space Centre SLV Launch Pad - Wikipedia, the free encyclopedia](#)

Indian space programme Indian Space **Research** Organisation (ISRO) Organisations Department of Space

BM25:
4/10

SIMBM25:
6/10

SIMBM25 Result

The Crawlers

☐ TFIDF ☐ BM25 ☒ SIMBM25

63 results found.

[Liquid Propulsion Systems Centre - Wikipedia, the free encyclopedia](#)

The Liquid Propulsion Systems Centre (LPSC; Hindi: द्रव नोदन प्रणाली केंद्र) is a **research**

[PSLV-C34 - Wikipedia, the free encyclopedia](#)

a new record of deploying the maximum number of satellites by Indian Space **Research** Organisation

[Sat Bir Singh Khalsa - Wikipedia, the free encyclopedia](#)

Massachusetts. Since 2007, Sat Bir Singh Khalsa has served as the Director of **Research** at both

[B. P. Loughridge - Wikipedia, the free encyclopedia](#)

at the Oklahoma Medical **Research** Foundation, and working as an exterminator for a pest control company. He

[Privadarshini Engineering College - Wikipedia, the free encyclopedia](#)

ensuring a high quality teaching learning and **research** environment. Motivated faculty - student

[Satish Dhawan Space Centre SLV Launch Pad - Wikipedia, the free encyclopedia](#)

Indian space programme Indian Space **Research** Organisation (ISRO) Organisations Department of Space

[PSLV-D1 - Wikipedia, the free encyclopedia](#)

and the mission was a total failure.[1][4] See also[edit] Indian Space **Research** Organisation Polar

[VIT University Jaipur - Wikipedia, the free encyclopedia](#)

2.5 M.Sc. programmes 2.6 The University also offers MBA programme 2.7 **Research** Programmes 3

[Johannes Kepler University Linz - Wikipedia, the free encyclopedia](#)

as well as web and the law. It offers those with academic background in various fields a **research**-led

[Sylhet International University - Wikipedia, the free encyclopedia](#)

institution of higher learning and **research** is dedicated to providing quality education commensurate

Query: Film and Actor

The Crawlers

☐ TFIDF ☒ BM25 ☐ SIMBM25

385 results found.

[Shah Rukh Khan filmography - Wikipedia, the free encyclopedia](#)

As an **actor** 2.2 As a host 3 See also 4 Footnotes 5 References 6 Bibliography 7 External links **Film**

[Jackie Shroff - Wikipedia, the free encyclopedia](#)

] Residence Mumbai, Maharashtra, India Nationality Indian Other names Jackie Shroff Occupation **Actor**

[Aamir Khan filmography - Wikipedia, the free encyclopedia](#)

for Best **Film** Filmfare Award for Best **Actor** Nominated—Academy Award for Best Foreign Language **Film**

[42nd Saturn Awards - Wikipedia, the free encyclopedia](#)

Site Burbank, California, U.S. Hosted by John Barrowman Highlights Most awards **Film**: Star Wars

[Aditya Roy Kapur - Wikipedia, the free encyclopedia](#)

) Aditya Roy Kapur (pronounced [ədɪˈtʏəˈ royˈ kəˈ pur]; born 16 November 1985) is an Indian **film actor**

[The Fighter - Wikipedia, the free encyclopedia](#)

Best Picture.[39] The **film** was nominated for six Golden Globe Awards: Best Picture (Drama), Best **Actor**

[Raj Kapoor \(Tamil film director\) - Wikipedia, the free encyclopedia](#)

director, **actor** Years active 1991–present Raj Kapoor is a Tamil **film** director who has directed

[Jonathan Banks - Wikipedia, the free encyclopedia](#)

actor. His first notable **film** roles were in the films Airplane!, 48 Hrs., and Beverly Hills Cop. He has

[Abhishek Bachchan filmography - Wikipedia, the free encyclopedia](#)

in 2011 Abhishek Bachchan is an Indian **actor** and producer known for his work in Bollywood films. He

[Zayed Khan - Wikipedia, the free encyclopedia](#)

(1980-07-05) 5 July 1980 (age 36) Mumbai, Maharashtra, India Nationality Indian Occupation **Actor**

BM25:
6/10

SIMBM25:
6/10

SIMBM25 Result

The Crawlers

☐ TFIDF ☐ BM25 ☒ SIMBM25

126 results found.

[Shah Rukh Khan filmography - Wikipedia, the free encyclopedia](#)

As an **actor** 2.2 As a host 3 See also 4 Footnotes 5 References 6 Bibliography 7 External links **Film**

[Aamir Khan filmography - Wikipedia, the free encyclopedia](#)

for Best **Film** Filmfare Award for Best **Actor** Nominated—Academy Award for Best Foreign Language **Film**

[Jackie Shroff - Wikipedia, the free encyclopedia](#)

] Residence Mumbai, Maharashtra, India Nationality Indian Other names Jackie Shroff Occupation **Actor**

[42nd Saturn Awards - Wikipedia, the free encyclopedia](#)

Site Burbank, California, U.S. Hosted by John Barrowman Highlights Most awards **Film**: Star Wars

[Aditya Roy Kapur - Wikipedia, the free encyclopedia](#)

) Aditya Roy Kapur (pronounced [ədɪˈtʏəˈ royˈ kəˈ pur]; born 16 November 1985) is an Indian **film actor**

[Abhishek Bachchan filmography - Wikipedia, the free encyclopedia](#)

in 2011 Abhishek Bachchan is an Indian **actor** and producer known for his work in Bollywood films. He

[Dynamite \(2015 film\) - Wikipedia, the free encyclopedia](#)

Dynamite (2015 **film**) - Wikipedia, the free encyclopedia Dynamite (2015 **film**) From Wikipedia

[List of roles and awards of Arshad Warsi - Wikipedia, the free encyclopedia](#)

for Best **Actor** in a Comic Role 2004 Muana Bhai M.B.B.S. Won [72] Apsara **Film** Producers Guild Award for Best

[87th Academy Awards - Wikipedia, the free encyclopedia](#)

Meron and Craig Zadan and directed by Hamish Hamilton,[8][9] **Actor** Neil Patrick Harris hosted

[Frozen \(2013 film\) - Wikipedia, the free encyclopedia](#)

Frozen (2013 **film**) - Wikipedia, the free encyclopedia Frozen (2013 **film**) From Wikipedia, the free

[Christopher Abbott - Wikipedia, the free encyclopedia](#)

Christopher Jacob Abbott (born February 1986)[1][2] is an American **actor**. He made his feature **film**

Query: Earth

BM25 Result

The Crawlers

☐ TFIDF ☒ BM25 ☐ SIMBM25

171 results found.

[Gone to Earth \(David Sylvian album\) - Wikipedia, the free encyclopedia](#)

Gone to **Earth** (David Sylvian album) - Wikipedia, the free encyclopedia Gone to **Earth** (David Sylvian

[List of science fiction sitcoms - Wikipedia, the free encyclopedia](#)

with a family on **Earth** Aliens in the Family – Jim Henson Productions' short-lived cross between

[Venus - Wikipedia, the free encyclopedia](#)

is the second planet from the Sun, orbiting it every 224.7 **Earth** days.[14] It has the longest rotation

[Jean Richer - Wikipedia, the free encyclopedia](#)

data-based estimate of the distance between **Earth** and Mars, which they then used to calculate

[PSLV-C34 - Wikipedia, the free encyclopedia](#)

including the primary payload Cartosat-2C. Cartosat-2C was placed in low **Earth** orbit at 9:44 IST

[David Wolfe \(nutritionist\) - Wikipedia, the free encyclopedia](#)

, author, Flat **Earth** theorist, promoter of alternative medicine,[1][2][3][4] and spokesman

[Gone to Earth \(Barclay James Harvest album\) - Wikipedia, the free encyclopedia](#)

Gone to **Earth** (Barclay James Harvest album) - Wikipedia, the free encyclopedia Gone to **Earth**

[Ed Lu - Wikipedia, the free encyclopedia](#)

, dedicated to protecting the **Earth** from asteroid strikes, later serving as its Chair. As of 2014 he

[Carolyn Porco - Wikipedia, the free encyclopedia](#)

rings and Enceladus; The Day the **Earth** Smiled; Science communicator & public speaker; Film consultant

[Continental drift - Wikipedia, the free encyclopedia](#)

that the earth's interior was fluid, and disagreeing with Lord Kelvin on the age of the **earth** [21] For example

BM25:
4/10

SIMBM25:
7/10

SIMBM25 Result

The Crawlers

☐ TFIDF ☒ BM25 ☐ SIMBM25

66 results found.

[Gone to Earth \(David Sylvian album\) - Wikipedia, the free encyclopedia](#)

Gone to **Earth** (David Sylvian album) - Wikipedia, the free encyclopedia Gone to **Earth** (David Sylvian

[List of science fiction sitcoms - Wikipedia, the free encyclopedia](#)

with a family on **Earth** Aliens in the Family – Jim Henson Productions' short-lived cross between

[Venus - Wikipedia, the free encyclopedia](#)

is the second planet from the Sun, orbiting it every 224.7 **Earth** days.[14] It has the longest rotation

[Jean Richer - Wikipedia, the free encyclopedia](#)

data-based estimate of the distance between **Earth** and Mars, which they then used to calculate

[PSLV-C34 - Wikipedia, the free encyclopedia](#)

including the primary payload Cartosat-2C. Cartosat-2C was placed in low **Earth** orbit at 9:44 IST

[Science fiction on television - Wikipedia, the free encyclopedia](#)

a scientist who finds the map of a cave that leads to the center of the **earth** and Trenreis naar de Toekomst

[List of science fiction television programs - Wikipedia, the free encyclopedia](#)

Apocalyptic and post-apocalyptic Biopunk Climate fiction Comedy Sitcoms Cyberpunk Derivatives Dying **Earth**

[PSLV-D1 - Wikipedia, the free encyclopedia](#)

fell back to **earth** and crashed in the Bay of Bengal. The satellite IRS-1E could not be placed in orbit

[Geosynchronous Satellite Launch Vehicle - Wikipedia, the free encyclopedia](#)

. The GSLV can place approximately 5,000 kg (11,000 lb) into an easterly low **Earth** orbit

[Liquid Propulsion Systems Centre - Wikipedia, the free encyclopedia](#)

[edit] This unit serves as LPSC headquarters, and is involved in research and development of **Earth**

[Titan 34D - Wikipedia, the free encyclopedia](#)

had a payload capacity of 14,515 kg (32,000 lb) to low **Earth** orbit, and of 5,000 kg (11,000 lb

Query: Television Series

BM25 Result

The Crawlers

☐ TFIDF ☒ BM25 ☐ SIMBM25

605 results found.

[List of science fiction television programs - Wikipedia, the free encyclopedia](#)

serial **series** short show special spin-off unaired episode/pilot. **Television** program types: animated

[Science fiction on television - Wikipedia, the free encyclopedia](#)

See also: Anime Robot characters from the Japanese science fiction **television series** Ganbare

[42nd Saturn Awards - Wikipedia, the free encyclopedia](#)

heavy changes in the **Television** categories, with all "Best TV **series**" categories being replaced

[Columbo - Wikipedia, the free encyclopedia](#)

is about the **television series**. For a list of episodes, see List of Columbo episodes. For other uses, see

[Amy Poehler - Wikipedia, the free encyclopedia](#)

member on the NBC **television series** Saturday Night Live from 2001 to 2008, Poehler became co-anchor

[Paper Moon \(TV series\) - Wikipedia, the free encyclopedia](#)

is her father.[2] Episodes[edit] Produced by Paramount **Television**, the **series** aired at 8:30 p.m.

[Entertainment Weekly - Wikipedia, the free encyclopedia](#)

-for-**television** films and new **television** program or **series**, as well as some **television** specials

[Jonathan Banks - Wikipedia, the free encyclopedia](#)

received critical acclaim for his role as Mike Ehrmantraut in the **television series** Breaking Bad

[The Nightmare Room - Wikipedia, the free encyclopedia](#)

, the **series** resembled the **television series** The Twilight Zone with teens taking the role as the main

[List of sitcoms broadcast in first-run syndication - Wikipedia, the free encyclopedia](#)

television series which ran from October 7, 1953 to July 2, 1954 on the DuMont **Television** Network

SIMBM25 Result

The Crawlers

☐ TFIDF ☐ BM25 ☒ SIMBM25

17 results found.

[List of science fiction television programs - Wikipedia, the free encyclopedia](#)

serial **series** short show special spin-off unaired episode/pilot. **Television** program types: animated

[Science fiction on television - Wikipedia, the free encyclopedia](#)

See also: Anime Robot characters from the Japanese science fiction **television series** Ganbare

[List of science fiction sitcoms - Wikipedia, the free encyclopedia](#)

fiction sitcom Serial Situation comedies Soap opera Sports Teen drama School **television series** Teen

[42nd Saturn Awards - Wikipedia, the free encyclopedia](#)

heavy changes in the **Television** categories, with all "Best TV **series**" categories being replaced

[Columbo - Wikipedia, the free encyclopedia](#)

is about the **television series**. For a list of episodes, see List of Columbo episodes. For other uses, see

[Amy Poehler - Wikipedia, the free encyclopedia](#)

member on the NBC **television series** Saturday Night Live from 2001 to 2008, Poehler became co-anchor

[The Sopranos \(season 5\) - Wikipedia, the free encyclopedia](#)

Tony custody of A.J.; after a **series** of infractions, Tony handles the Feech La Manna problem

[Mad Men \(season 3\) - Wikipedia, the free encyclopedia](#)

of Mad Men episodes The third season of the American **television** drama **series** Mad Men premiered

[The West Wing - Wikipedia, the free encyclopedia](#)

drama **television series** created by Aaron Sorkin that was originally broadcast on NBC from September

[All in the Family - Wikipedia, the free encyclopedia](#)

States as one of the greatest **television series** of all time.[3] The show ranked number-one

BM25:

5/10

SIMBM25:

8/10

Conclusion

- SIMBM25 scoring function is an extension of BM25.
- It uses structural contextual information to improve results.
- SIMBM25 gives better precision than BM25. In other words, SIMBM25 is at least as good as BM25.
- Quality of SIMBM25 depends upon top k results of BM25.
- Further scoring can be improved by including similarity of the results and query.
- A graph can be built taking nodes as {query, res1, res2,.....,resN} and SimRank can be applied to get the similarity value.

Thank You