

Vertex Page Clustering

Kaio Cesar Alves Reis, Matteo De Santis, Francesco Forgione

- Sviluppo e implementazione dell'algoritmo di clustering di Vertex
- Dato un sito Web, l'algoritmo deve essere in grado di raggruppare le varie pagine del dominio in cluster
- Ciascun cluster fa riferimento ad un'entità
- L'algoritmo deve essere in grado di riconoscere l'entità descritta nella pagina Web

- L'idea di base di Vertex è quella di valutare le pagine Web in base alla loro struttura
- Per fare questo l'algoritmo viene suddiviso in tre passi

- Nel primo passo viene implementata una hash table che contiene tutti i **masked shingle vector**, e il relativo punteggio, di tutte le pagine del dominio in esame

Input: Sample pages P from Web site to be clustered;

Output: Set of clusters;

/ First Pass */*

Initialize hash table H to empty;

for each page $p \in P$ **do**

 Let v be the shingle vector for p ;

for each 6/8, 7/8, and 8/8 masked shingle vector v' covering v **do**

if v' is in hash table H **then**

 Increment the count for v' ;

else

 Insert v' with count 1 into H ;

end if

end for

end for

- Nel secondo passo viene creato un set contenente tutti i masked shingle vector di ogni pagina con punteggio maggiore, i quali coprono gli shingle vector delle pagine campione

/ Second Pass */*

for each 8/8 vector v in H in increasing order of counts
do

Let v' be the masked shingle vector in H with maximum count covering v ;

Decrement counts of all masked shingle vectors $\neq v'$ in H covering v (by v 's count);

end for

Delete masked shingle vectors with counts less than threshold from H ;

- Nel terzo passo, per ogni pagina del dominio vengono confrontati i masked shingle vector e successivamente questa viene assegnata ad un cluster

/ Third Pass */*

For each masked shingle vector $v \in H$, $C_v = \emptyset$;

for each page $p \in P$ with shingle vector v **do**

 Let v' be the masked shingle vector in H with maximum count covering v ;


 Add p to $C_{v'}$;

end for


return $\{(C_v, v, \text{count for } v) : v \in H\}$;

- Estrazione della ground truth relativa al dominio
- Per ciascuna pagina Web del dominio vengono estratti tutti i tag html
- I tag vengono utilizzati per creare tutti gli shingle di lunghezza 10
- Viene associata una firma ad ogni pagina Web creata in base a tutti gli shingle di questa
- Viene creato un file contenente pagine e relative firme

Input

- 
- Il file precedentemente creato viene usato come input per l'algoritmo di clustering

Output

- L'output del programma è un file contenente tutti i cluster individuati
- 

Metrice: Guide2Research

Guide2Research

IEEE Communications Surveys and Tutorials

Posted on November 4, 2020 in Journal

Journal Ranking & Metrics

Impact Score : 20.3
G2R H-Index : 62
JCR Impact Factor : 23.7
Scopus Citescore : 52.6
SCIMAGO SJR : 7.553
SCIMAGO H-index : 169

Guide2Research

Journal Info

IWCMC 2021 : International Wireless Communications and Mobile Computing Conference

Posted on July 20, 2020 in Conferences

Conference Information

Submission Deadline : Sunday 10 Jan 2021
Conference Dates : Jun 28, 2021 - Jul 2, 2021
Conference Address : Beijing, China
Conference & Submission Link : <https://iwcmc.org/2021/>

Proceedings indexed by :


Conference Organizers : ([Deadline extended ? Click here to edit](#))

Conference Ranking & Metrics (This is a TOP Conference)

Impact Score : 3.96
#Contributing Top Scientists : 50
Guide2Research Overall Ranking : 163
Category Rankings :
Networks and Communications : 46

All Conferences

Top Conferences

Top 1000 Journals

Top 1000 Scientists

University Rankings

Special Issues

Guide2Research



Herbert Simon

Carnegie Mellon University
United States

G2R World Ranking : 2nd
G2R United States Ranking : 2nd

H-Index & Metrics

Google H-Index : 182
Number of Google Citations : 379,912
Number of Articles on DBLP : 80

Awards & Achievements

Guide2Research

University Ranking for Computer Science:

The University Ranking for Computer Science & Electronics was prepared by Guide2Research, one of the leading portals for computer science research providing trusted data on scientific contributions since 2014.

The ranking represents h-index, and DBLP values gathered by August 25th 2020. It was based on a detailed examination of more than 6000 computer scientists' profiles on Google Scholar and DBLP.

[Read More...](#)

Search by name :

View by country :

1-100			101-200		201-300		301-400		401-500		501-600		601-700		
World			National			University			#Scholars			Σ DBLP		Σ H-Index	
1			1			Carnegie Mellon University United States			116			22,440		7152	
2			2			MIT United States			83			15,771		5859	

All Conferences

Top Conferences

Top 1000 Journals

Top 1000 Scientists

University Rankings

Special Issues

Research Blog

All Conferences

Top Conferences

Top 1000 Journals

Top 1000 Scientists

University Rankings

Special Issues

Research Blog

Contact us

Cluster	Precision	Recall	F1
Conference	0,96	1	0,98
Journal	1	1	1
Special Issue	0,79	1	0,88
User	1	1	1

Metrice: The Movie DB

The screenshot shows the TMDb profile for Tom Hanks. The header includes the TMDb logo and navigation links for Movies, TV Shows, People, and More. Below the header, there's a tab bar with Overview, Media, Fandom, and Share. The main content area features a large portrait of Tom Hanks on the left and a biography on the right. The biography text reads: "Thomas Jeffrey Hanks (born July 9, 1956) is an American actor and filmmaker. Known for both his comedic and dramatic roles, Hanks is one of the most popular and recognizable film stars worldwide, and is widely regarded as an American cultural icon." Below this, it mentions his breakthrough roles in Splash (1984) and Big (1988), and his Academy Awards for Best Actor for starring as a gay lawyer suffering from AIDS in Philadelphia (1993) and a young man with below-average IQ in Forrest Gump (1994). It also lists his collaborations with Steven Spielberg on five films: Saving Private Ryan (1998), Catch Me If You Can (2002), The Terminal (2004), Bridge of Spies (2015), and The Post (2017), as well as the 2001 miniseries Band of Brothers, which launched him as a director, producer, and screenwriter.

The screenshot shows the TMDb page for the movie Interstellar (2014). The header includes the TMDb logo and navigation links for Movies, TV Shows, People, and More. Below the header, there's a tab bar with Overview, Media, Fandom, and Share. The main content area features a large movie poster on the left and a detailed overview on the right. The overview text reads: "Mankind was born on Earth. It was never meant to die here." Below this, it mentions the adventures of a group of explorers who make use of a newly discovered wormhole to surpass the limitations on human space travel and conquer the vast distances involved in an interstellar voyage. The director and writer are listed as Christopher Nolan, and the writer is Jonathan Nolan. The movie is marked as "Now Streaming" with a green icon.

The screenshot shows the TMDb page for the TV show The Mandalorian (2019). The header includes the TMDb logo and navigation links for Movies, TV Shows, People, and More. Below the header, there's a tab bar with Overview, Media, Fandom, and Share. The main content area features a large show poster on the left and a detailed overview on the right. The overview text reads: "Bounty hunting is a complicated profession." Below this, it mentions the fall of the Galactic Empire, lawlessness has spread throughout the galaxy, and a lone gunfighter makes his way through the outer reaches, earning his keep as a bounty hunter. The creator is listed as Jon Favreau. The show is marked as "Now Streaming Watch Now" with a green icon.

Let's Chat

Discuss → Movies

Subject	Status	Replies	Last Reply
Very good. Promising Young Woman	Open	48	Jan 16, 2021 at 10:55 AM by MangoLloyd
How should Marvel approach doing this? Fantastic Four	Open	21	Jan 16, 2021 at 7:20 AM by Damienracer

Cluster	Precision	Recall	F1
Discover	1	0,74	0,84
Movie	0,78	0,93	0,90
Network	0,94	1	0,98
Person	0,96	1	0,97
Review	0,76	1	0,88
Talk	0,43	0,9	0,71
Tv	0,82	0,92	0,78
U	0,96	0,99	0,96

Study Innovation Management: All you need to know

Author: **Megan Thudium** — Last update: 27 August 2020

Innovation Management is a versatile and exciting degree to study at university. It's the perfect subject if you are forward-thinking and creative and enjoy combining high-level strategy with hands-on work on new business ideas. Students that pursue this field are well-equipped to become the innovators, the thinkers, the creatives, and the leaders of tomorrow. Read on to learn more:

What is Innovation Management?

Innovation Management is a field promoting organizational innovation and includes processes that encourage an organization to strategically innovate. Studying Innovation Management at university will equip you to solve the common challenges organizations face to successfully innovate, whether that company is a corporation, enterprise, start-up, or small business.





A university degree in Innovation Management will provide you with insight into the role that innovation plays in organizations, how it ties into corporate strategy, product development, market positioning, and new business creation.

Innovation Managers perform tasks like

- Promote innovation and innovative processes throughout an org;
- Develop and collect ideas for innovation,
- Work directly with leadership to make strategic long-term plans.

As a graduate of Innovation Management, you can fundamentally transform and company cultures. Ready to make an impact? If yes, then a university is your choice!

FEATURED UNIVERSITIES:

	Falmouth University Falmouth, United Kingdom 70 study programmes
	ECAM Lyon Lyon, France 4 study programmes
	Hertie School Berlin, Germany 4 study programmes
	CODE University of Applied Sciences Berlin, Germany 2 study programmes

Italy

All programmes in Italy

Italy's historic treasures and magnificent cuisine continue to draw visitors as they have done for centuries. From the snow-capped Dolomites in the North to the warm maritime Mediterranean culture of the South, Italy's regions are diverse and exciting. For the international student seeking world-class but affordable education, with a wide range of social and leisure activities, few countries are as exciting a choice as Italy.

The country's cultural importance extends back to ancient times and the past greets the student of history and archaeology on every street corner. For many, Rome is still the Eternal City, the centre of an empire that – some 2,000 years ago – stretched from rainy Britannia to the deserts of Syria. At its height it also encompassed Egypt, much of North Africa and nearly all of Continental Europe.

Italy's Ancient Roman monuments, magnificent as they are, tell only part of the story. The Renaissance of the 14th to 17th century originated in Italy (as the *Rinascimento*) and sparked innovation and revolutions in art, architecture, philosophy, religion and the sciences across Europe. Intellectual activity and creativity flourished, carrying gems of ancient and medieval thought into the modern world.

Studying at a university in Italy

It's perhaps not surprising that many international students feel instantly at home in Italy. The University of Bologna has been welcoming students since 1088 – it is the oldest university in the world and its foundation gave the city one of its nicknames: *la dotta*, "the learned one". Bologna remains at the leading edge of academic innovation in the 21st century. The Bologna Process was initiated in the city, and has established a set of standards for excellence in higher education across the European Higher Education Area (EHEA).

Italian universities take pride in their heritage and rankings. The University of Rome Tor Vergata offers more than a dozen international degree courses that are taught in English. Founded only in 1982, Tor Vergata was ranked one of

University of Kent


The University is one of the country's leading academic institutions producing world-class research, rated internationally excellent and leading the way in many fields of study.

We are a forward-thinking research institution, committed to the transformative power of education and research and to the development and support of our students and staff.


— kent.ac.uk



Bachelor's programmes Master's programmes Doctorate programmes

Bachelors at the University of Kent

	
Population:	60,918,000
Currency:	EUR
Listed universities:	57
Study programmes:	419
Ranked universities:	52
Country Rank (2018):	9
Look at study programmes in Italy	

FEATURED UNIVERSITY:

	Ca' Foscari University of Venice Venice, Italy 20 study programmes
---	--

	
University of Kent	
Location:	Canterbury (United Kingdom)
Founded:	1965
Students:	19,300 (17% international)
Academic staff:	2,000
Ranking: 	#170 (Europe) #351 (World)
Listed study programmes:	521
Website:	kent.ac.uk

Cluster	Precision	Recall	F1
Faculty	0,92	1	0,95
Search	0,87	1	0,93
University	0,54	1	0,70

- Il raggruppamento delle pagine nel dataset è stato fatto manualmente
- L'algoritmo risulta avere una complessità n^2
- Processa e classifica circa 15 pagine al secondo
- L'esperimento è stato eseguito con una threshold pari a 10
- Con valori di threshold maggiori il numero di cluster individuati è minore