



Promoting **Serendipitous Discovery** of Academic Literature with Transformers and Visual Analytics



Arpit **Narechania**



Alireza **Karduni**



Ryan **Wesslen**



Emily **Wall**





Motivation / Problem: Story cognitive biases

Especially a problem for interdisciplinary studies (e.g., different disciplines use same words differently)

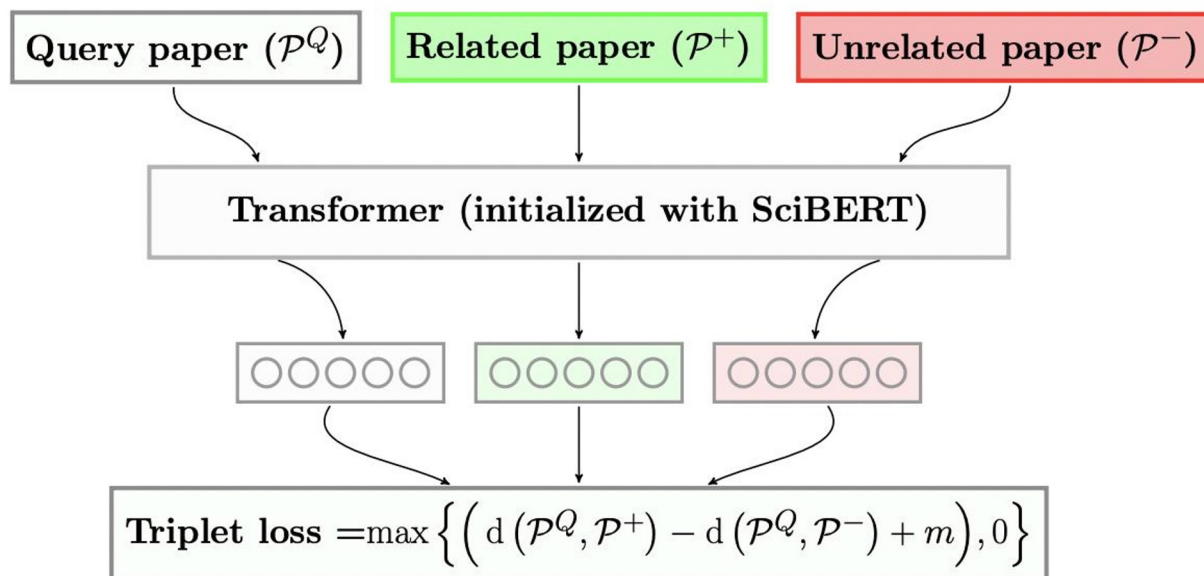
Transformers/Contextual embeddings

SPECTER: Document-level Representation Learning using Citation-informed Transformers

Arman Cohan^{†*} Sergey Feldman^{†*} Iz Beltagy[†] Doug Downey[†] Daniel S. Weld^{†,‡}

[†]Allen Institute for Artificial Intelligence

[‡]Paul G. Allen School of Computer Science & Engineering, University of Washington



SPECTER **outperformed** nine (9) benchmarks (e.g., doc2vec, fasttext, SIF, ELMo, SciBERT, Sent-BERT) on seven (7) NLP tasks (SciDocs):

- **Document classification** (topic or subject header)
- **User activity prediction** (co-view or co-read)
- **Citation prediction** (cite or co-cite)
- **Recommendation**

Formative Study



Formative Study



DG1. Enable **serendipitous identification of semantically related articles** that do not necessarily have shared keywords through visual exploration.

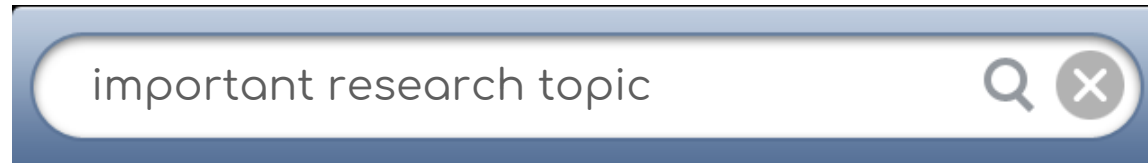


Formative Study



DG1. Enable **serendipitous identification of semantically related articles** that do not necessarily have shared keywords through visual exploration.

DG2. Facilitate a **familiar search functionality** to what users are currently accustomed to, such as keyword and author search.



Formative Study

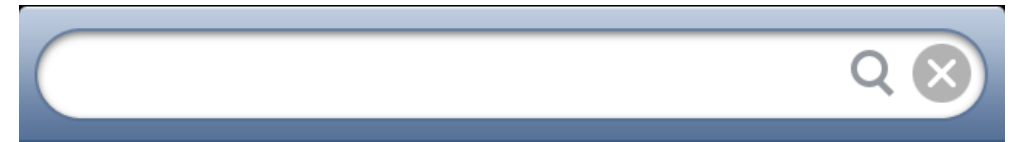


DG1. Enable **serendipitous identification of semantically related articles** that do not necessarily have shared keywords through visual exploration.

DG2. Facilitate a **familiar search functionality** to what users are currently accustomed to, such as keyword and author search.

DG3. Afford users to find semantically related articles by **search based on the author's own ideas** in the form of unpublished sentences / abstracts.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi sed scelerisque nisl. Donec venenatis odio nec lectus pellentesque, in suscipit mi pharetra. Curabitur et condimentum nisi. Duis fringilla elementum ante, dictum rutrum elit tempus et. Pellentesque eget dapibus turpis. Phasellus a euismod justo, sit amet porta ante. Curabitur euismod sed neque vel consequat. Fusce a tristique justo. Ut dignissim lobortis luctus. Nullam nec scelerisque dolor. Nullam a mollis ipsum. Pellentesque ullamcorper felis at purus convallis commodo.



Formative Study

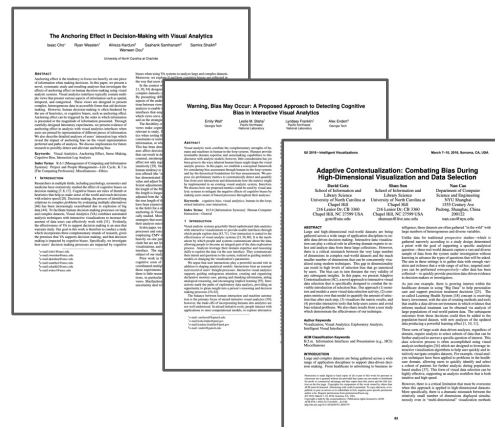


DG1. Enable **serendipitous identification of semantically related articles** that do not necessarily have shared keywords through visual exploration.

DG2. Facilitate a **familiar search functionality** to what users are currently accustomed to, such as keyword and author search.

DG3. Afford users to find semantically related articles by **search based on the author's own ideas** in the form of unpublished sentences / abstracts.

DG4. Enable users to interact with a **visual overview** of a group of papers.



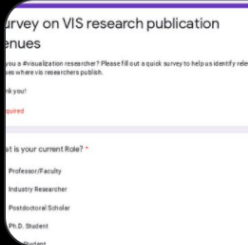
Where do VIS researchers publish?

 **Arpit Narechania**
@arpitnarechania

Are you a [#datavisualization](#) researcher? Please fill out a quick 5-min survey to help us identify relevant venues where vis researchers publish (e.g., IEEE VIS, ACM CHI).

Survey 📄 forms.gle/pg8bJHa2BP3sfE...

Also, Please RT 🙏.



Survey on VIS research publication venues

Are you a #visualization researcher? Please fill out a quick survey to help us identify relevant venues where vis ...

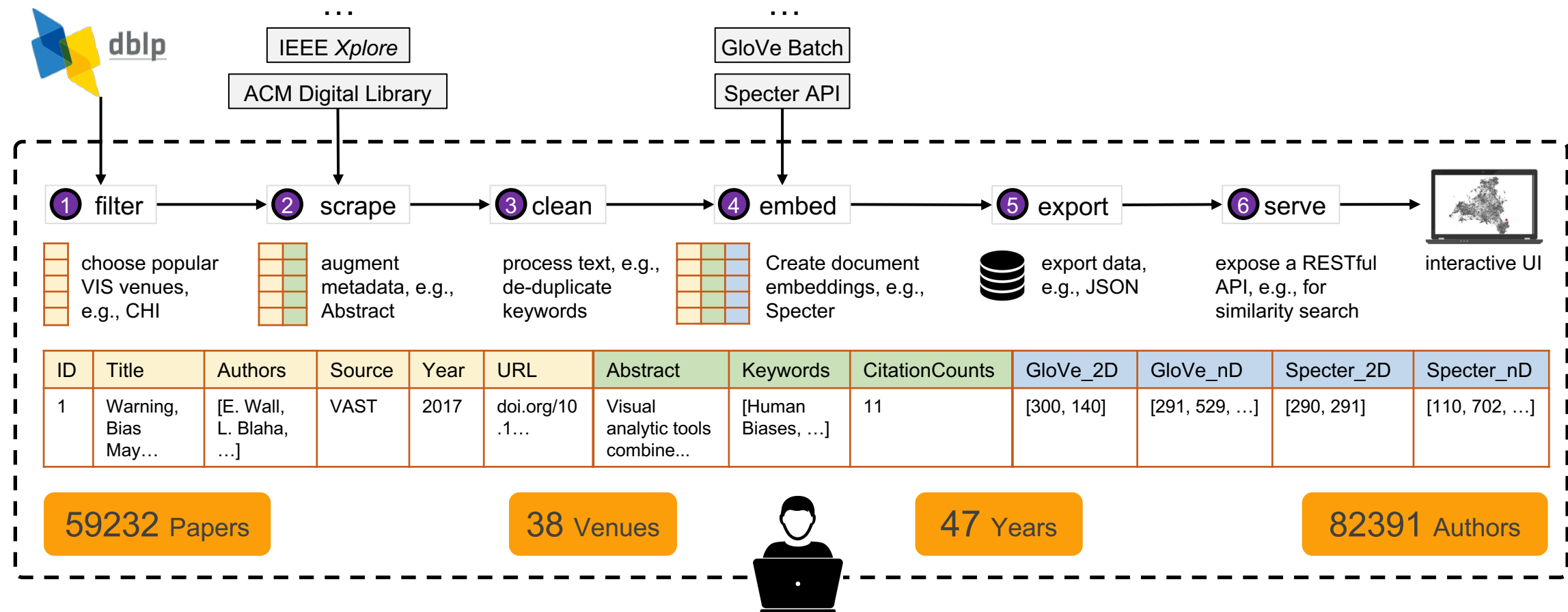
docs.google.com

1:34 PM · Dec 17, 2020 · Twitter Web App

twitter.com/arpitnarechania/status/1339640045754527749



System Architecture



Dataset



[Browse](#)

VitaLITy: A Dataset of Academic Articles

[Cite](#) [Download all \(1.79 GB\)](#) [Share](#) [Embed](#) [+ Collect](#) [...](#)

Dataset posted on 30.03.2021, 03:30 by [Arpit Narechania](#), [Alireza Karduni](#), [Ryan Wesslen](#), [Emily Wall](#)

A dataset of scraped metadata from **59,232** academic articles including paper titles, keywords, abstracts, citation counts, year of publication, and word embeddings (GloVe and Specter) from **38** popular venues for visualization research.

https://figshare.com/articles/dataset/VitaLITy_A_Dataset_of_Academic_Articles/14329151

System

vitaLITy

Specter Saved Papers (3)

Showing 59232/59232

Search

Column All All All

Info	ID	Title	Authors	Source	Year	Abstract	Keywords	CitationCo	Map	Selec	Save

System | Search by a List of Seed Papers



By Papers (8) By Abstract Output Similar (25)

Dimensions nD Count 25 Find Similar Papers

Showing 8/8 Search Column + All All All

Info	Title	Map	Save	Delete
	Investigating Effects of Visual Anchors on Decision-Making about			
	Toward a Design Space for Mitigating Cognitive Bias in Vis.			
	Mitigating the Attraction Effect with Visualizations.			
	Warning, Bias May Occur: A Proposed Approach to Detecting Cog			
	A Task-Based Taxonomy of Cognitive Biases for Information Visu			
	The Attraction Effect in Information Visualization.			

By Papers (8) By Abstract Output Similar (25)

Showing 25/25 Search Column + All All All

Info	Title	Sim	Map	Sele	Save
		0.062300.4355			
	A Formative Study of Interactive Bias Metrics in Visu	0.4355			
	Studying Biases in Visualization Research: Framework	0.3437			
	So, What Are Cognitive Biases?	0.3283			
	CogTool-Explorer: a model of goal-directed user expl	0.2737			
	Just the other side of the coin?: from error- to insight	0.2593			
	Working memory load as a novel tool for evaluating	0.1873			

System | Search by a WIP Abstract



By Papers (0) By Abstract **Output Similar (0)**

Count 25 Find Similar Papers

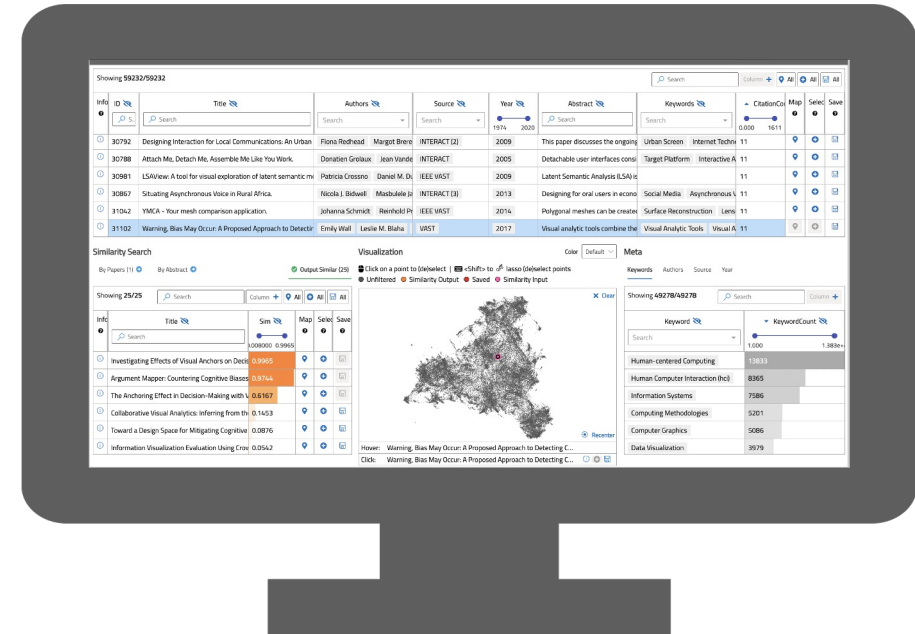
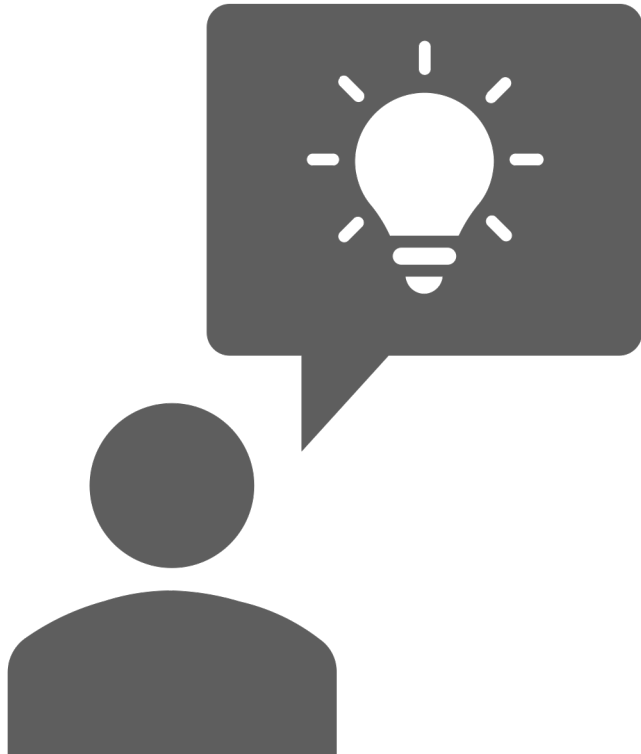
VITALITY: Promoting Serendipitous Discovery of Academic Literature with Transformers & Vi...

There are a few prominent practices for conducting reviews of academic literature, including searching for specific keywords on Google Scholar or checking citations from some initial seed paper(s). These approaches serve a critical purpose for academic literature reviews, yet there remain challenges in identifying relevant literature when similar work may utilize different terminology (e.g., mixed-initiative visual analytics papers may not use the same terminology as papers on model-steering, yet the two topics are relevant to one another). In this paper, we introduce a system, VitaLITy, intended to complement existing practices. In particular, VitaLITy promotes serendipitous discovery of relevant literature using transformer language models, allowing users to find semantically similar papers in a word embedding space given (1) a list of input paper(s) or (2) a working abstract. VitaLITy visualizes this document-level embedding space in an interactive 2-D scatterplot using dimension reduction. VitaLITy also summarizes meta information about the document corpus or search query, including keywords and co-authors, and allows users to save and export papers for use in a literature review. We present qualitative findings from an evaluation of \system, suggesting it can be a promising complementary technique for conducting academic literature reviews. Furthermore, we contribute data from 37 popular data visualization publication venues in VitaLITy, and we provide scrapers for the open-source community to continue to grow the list of supported venues.

By Papers (0) By Abstract **Output Similar (25)**

Showing 25/25 Search Column All All All

Info	Title	Sim	Map	Select	Save
	Search	0.0001.000			
	PaperVis: Literature Review Made Easy.	1			
	VisualBib: Narrative Views for Customized Bibli	0.4249			
	EvolutionWorks - Towards Improved Visualizati	0.3083			
	TopicPanorama: A Full Picture of Relevant Topic	0.2755			
	CiteRivers: Visual Analytics of Citation Patterns	0.1929			
	Serendip: Topic model-driven visual exploration	0.1906			



supporting serendipitous discovery



custom scrapers

Google
scholar



☐ Articles ☒ include patents ☐ Case law

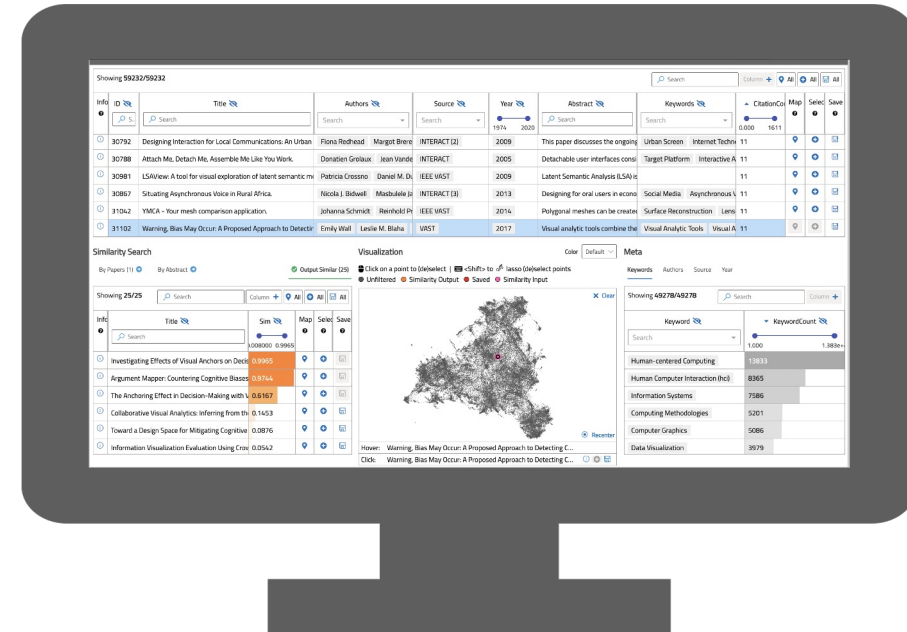


snowballing references

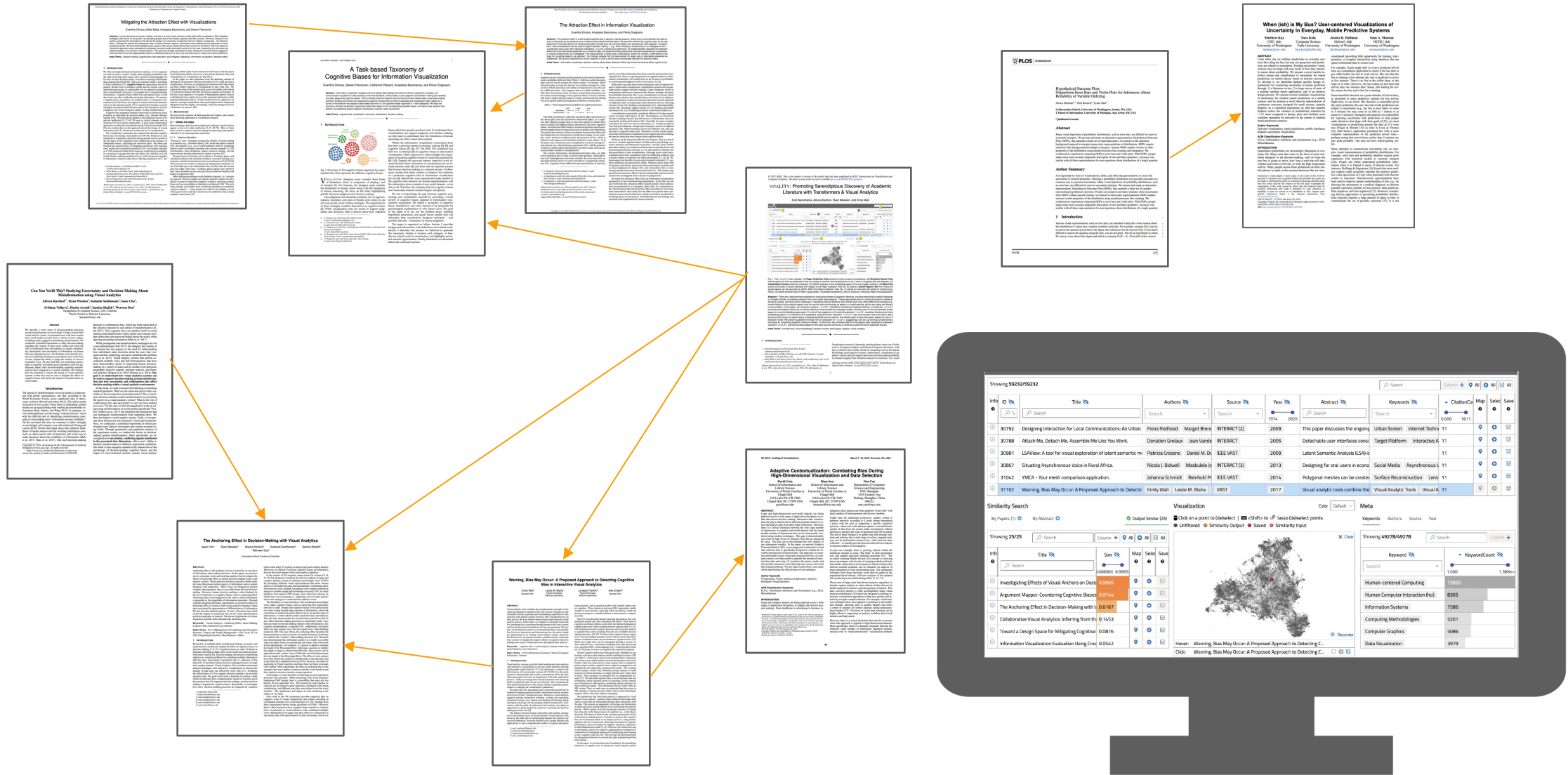


dblp

computer science bibliography



part of a larger workflow



improve model + add citation data



vitaLity

Specter Saved Papers (3)

Showing 59232/59232

Search

Column + All All All

Info	ID	Title	Authors	Source	Year	Abstract	Keywords	CitationCo	Map	Sele	Save
	30792	Designing Interaction for Local Communications: An Urban	Fiona Redhead Margot Brere	INTERACT (2)	2009	This paper discusses the ongoing	Urban Screen Internet Techn	11			
	30788	Attach Me, Detach Me, Assemble Me Like You Work.	Donatien Grolaux Jean Vande	INTERACT	2005	Detachable user interfaces consi	Target Platform Interactive A	11			
	30981	LSAView: A tool for visual exploration of latent semantic m	Patricia Crossno Daniel M. Du	IEEE VAST	2009	Latent Semantic Analysis (LSA) is		11			
	30867	Situating Asynchronous Voice in Rural Africa.	Nicola J. Bidwell Masbulele Ja	INTERACT (3)	2013	Designing for oral users in econo	Social Media Asynchronous v	11			
	31042	YMCA - Your mesh comparison application.	Johanna Schmidt Reinhold Pr	IEEE VAST	2014	Polygonal meshes can be create	Surface Reconstruction Lens	11			

<https://vitality-vis.github.io>

Info

Search

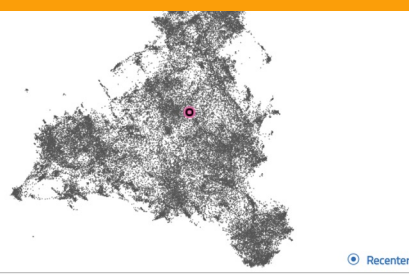
Sim 0.008000 0.9965

Map

Sele

Save

	Investigating Effects of Visual Anchors on Decis	0.9965			
	Argument Mapper: Countering Cognitive Biases	0.9744			
	The Anchoring Effect in Decision-Making with V	0.6167			
	Collaborative Visual Analytics: Inferring from th	0.1453			
	Toward a Design Space for Mitigating Cognitive	0.0876			
	Information Visualization Evaluation Using Crow	0.0542			



Recenter

Keyword

Search

KeywordCount 1.000 1.383e+

Human-centered Computing	13833
Human Computer Interaction (hci)	8365
Information Systems	7586
Computing Methodologies	5201
Computer Graphics	5086
Data Visualization	3979

Hover: Warning, Bias May Occur: A Proposed Approach to Detecting C...

Click: Warning, Bias May Occur: A Proposed Approach to Detecting C...



Qualitative Evaluation



Usage Scenarios

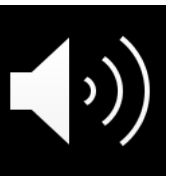


Dataset



Web Scrapers





Promoting **Serendipitous Discovery** of Academic Literature with Transformers and Visual Analytics



Arpit **Narechania**



Alireza **Karduni**



Ryan **Wesslen**



Emily **Wall**

