Academic Recommender

Implementation Comparison on Real Data

Implementation Comparison*

Method	Inference Time on University Data (secs)	Inference Time on Google Scholar (secs)
Mean Vector Cosine Similarity • Using nested averaging over research interests • Works both with words and phrases (sub-word mapping, cluster mapping)	Token-based: 7.50 Phrase-based: 4.05	N/A
Word Mover's Distance (token-based) • Using wmdist method • Using "fasttext-wiki-news-subwords-300" model • Works only with words, when all words are available in model vocab • Creates dictionary for each comparison pair, thus cannot be cached	12.47	N/A
Sentence Transformers • Using cosine similarity method • Using "all-MiniLM-L12-v2" model • Encodes research interests into single vector (tensor) • Creates encoding map for all data records, thus can be cached	One-time encoding: 395.65 1.40	One-time encoding: 44.13 0.26
BERT • Using cosine similarity • Using "bert-base-uncased" model • Encodes research interests into single vector (tensor) • Creates encoding map for all data records, thus can be cached	One-time encoding: 1003.1 2.69	One-time encoding: 130.17 2.01

* COMPARISON IS BASED ON THE SAME INPUT STUDENT, RECOMMENDING TOP 5 PROFESSORS. WE RAN ALL SCENARIOS 5 TIMES AND GOT THE AVERAGE RESULT.

Input student*	Sentence Transformers (all-MiniLM-L12-v2)	BERT (bert-base-uncased)
Name: Kathryn Morris **Research Interests:** • Mathematical Physics • Applied Mathematics • Geometry • Probability • Differential Equations • Algebra • Statistics • Topology • Number Theory • Calculus	**Best Match #1:** Bei Cai • Particle Astrophysics • Dark Matter • Neutrinos • Cosmic Rays • Physics and Engineering Education Research	**Best Match #1:** Hazhir Aliahmadi • Optimization • Systems Physics • Econophysics • Applied Fractional calculus
	Best Match #2: Felipe A. Gomes Ferreira • Theoretical Particle Physics • Cosmology • Artificial Intelligence • Data Science • AI Project Manager	**Best Match #2:** Francesco Cellarosi • Dynamics • Probability • Number Theory
	Best Match #3: Wulin Suo • Finance Derivatives • Risk Management • Asset Pricing • & Applied Mathematics	**Best Match #3:** Tyler Meadows • ODEs • Mathematical Biology • Dynamical Systems • Impulsive differential equations
	Best Match #4: Jaskirat Singh • Artificial Intelligence • Machine Learning • Deep Learning • Computer Vision • Data Science	**Best Match #4:** Andrew D. Lewis • Differential geometry • geometric mechanics • geometric control theory
	Best Match #5: Joseph Bramante • Particle Theory • Dark Matter • Cosmology • Theoretical Physics	**Best Match #5:** Felicia Maria G. Magpantay • differential equations • numerical analysis • mathematical biology

* COMPARISON IS BASED ON THE SAME INPUT STUDENT, RECOMMENDING TOP 5 PROFESSORS FROM GOOGLE SCHOLAR DATASET.

Input student	Sentence Transformers (all-MiniLM-L12-v2)	BERT (bert-base-uncased)
Name: Bethany Silva **Research Interests:** • Cloud Computing • Artificial Intelligence • Machine Learning • Human-Computer Interaction • Data Science • Blockchain • Cybersecurity • Computer Vision • Software Engineering • Quantum Computing	**Best Match #1:** Marwa A. Elsayed • Cybersecurity • Cloud Computing • Internet of Things • Applied Machine Learning • Software Engineering	**Best Match #1:** Nafiz Sadman • Natural Language Processing • Machine Learning • Deep Learning • Data Science • Cybersecurity
	Best Match #2: Jaskirat Singh • Artificial Intelligence • Machine Learning • Deep Learning • Computer Vision • Data Science	**Best Match #2:** Marwa A. Elsayed • Cybersecurity • Cloud Computing • Internet of Things • Applied Machine Learning • Software Engineering
	Best Match #3: Aman Anand • Data Science • Machine Learning • Deep Learning • IOT • Robotics	**Best Match #3:** Jaskirat Singh • Artificial Intelligence • Machine Learning • Deep Learning • Computer Vision • Data Science
	Best Match #4: Mohamed Sami Rakha • Software Engineering • Artificial Intelligence • Data Science • Cyber-security	**Best Match #4:** Yuanzhu Chen (陈远筑, 陳遠築) • Computer networking • mobile computing • complex networks • graph theory • web information retrieval
	Best Match #5: Vandad Davoodnia • Computer Vision • Computer Graphics • Machine Learning • Artificial Intelligence • Virtual Humans	**Best Match #5:** Farhana Zulkernine • Big data management • deep learning • cognitive computing • computer vision • IoT & text analytics

Input student	Sentence Transformers (all-MiniLM-L12-v2)	BERT (bert-base-uncased)
Name: Robert Young Jr. **Research Interests:** • Immunology • Marine Biology • Evolutionary Biology • Microbiology • Conservation Biology • Genetics • Ecology • Cell Biology • Biochemistry • Neuroscience	**Best Match #1:** Baharul Choudhury • Genomics • Evolution • Molecular Biology • Population Genetics • Conservation Biology	**Best Match #1:** Baharul Choudhury • Genomics • Evolution • Molecular Biology • Population Genetics • Conservation Biology
	Best Match #2: Natalya Odoardi • Microbiology • Immunology • Reproductive Immunology • Cancer Biology	**Best Match #2:** christopher g. eckert • Ecology • Evolution • population genetics • Botany • conservation biology
	Best Match #3: Peter T. Boag • conservation biology • environmental biology • Evolution • Genetics • ecology	**Best Match #3:** Natalya Odoardi • Microbiology • Immunology • Reproductive Immunology • Cancer Biology
	Best Match #4: christopher g. eckert • Ecology • Evolution • population genetics • Botany • conservation biology	**Best Match #4:** Peter T. Boag • conservation biology • environmental biology • Evolution • Genetics • ecology
	Best Match #5: Emily Červenka • Molecular biology • cell biology • Bioplastics • electromicrobiology	**Best Match #5:** Emily Červenka • Molecular biology • cell biology • Bioplastics • electromicrobiology

Input student	Sentence Transformers (all-MiniLM-L12-v2)	BERT (bert-base-uncased)
Name: Mary Williams **Research Interests:** • Rhetoric • Narrative Theory • Drama • Historiography • World Literature • Comparative Literature • Poetry • Novel • Literary Criticism • Creative Writing	**Best Match #1:** Rebecca Luce-Kapler • literary practices • Writing • Poetry • curriculum	**Best Match #1:** Sharday Mosurinjohn • Psychedelics • religious studies • Philosophy • cultural studies • contemporary art
	Best Match #2: Golam Rabbani • Baul Literature Music and Spirituality • Creative Industries • Ethnomusicology • Literary and Cultural Studies • Film and Media	**Best Match #2:** Golam Rabbani • Baul Literature Music and Spirituality • Creative Industries • Ethnomusicology • Literary and Cultural Studies • Film and Media
	Best Match #3: Claire Ahn • Multiliteracies • visual rhetoric • Genre • secondary English Language Arts • critical social justice education	**Best Match #4:** Trish Salah • trans studies • postcolonial feminism • Psychoanalysis • poetry and poetics • cultural production
	Best Match #4: Trish Salah • trans studies • postcolonial feminism • Psychoanalysis • poetry and poetics • cultural production	**Best Match #4:** Sean Rhoads • Ecocriticism • science fiction • Monsters • Film • Literature
	Best Match #5: Veronika Kratz • environmental humanities • US literature and culture • desert humanities • science studies	**Best Match #5:** Darshana Chakrabarty • Film & Media Studies • South Asian Studies • Gender and Sexuality Studies • Queer Theories • Shakespeare

Evaluation with Groundtruth

We consider example researcher from specific university (i.e. Queen's U) as input, and take his/her coauthors (& labmates with close interests) mentioned in papers indexed in Google Scholar as groundtruth for best matching researchers, provided that they are also from the same university.

While having limitations, we filter examples to have collaborators within the same research area, and have completed Google Scholar profiles. With all limits, this evaluation is feasible and worth-investigating.

As per evaluation metric, XXX is used.

Input Student	BERT results	Groundtruth
Name: Mahdiyar Molahasani **Research Interests:** • Continual Learning • Deep Learning • Super-Resolution • Neuromorphic Adv MRR (out of 1.0) = 0.20	**Best Match #1:** Pritam Sarkar • Deep Learning • Self-supervised Learning • Computer Vision • Multimodal • Time-series	**Best Match #1:** Ali Etemad • Artificial Intelligence • Deep Learning • Wearable Computing • Affective Computing • Human-Centred AI
	Best Match #2: Mojtaba Kolahdouzi • Deep learning • Face Recognition • Image Processing • Fair machine learning	**Best Match #2:** Michael Greenspan • Computer Vision
	Best Match #3: Tim von Hahn • machine learning • condition monitoring • predictive maintenance • advanced manufacturing • deep learning	**Best Match #3:** Haleh Damirchi • Machine Vision • Representation Learning
	Best Match #4: Debaditya Shome • Diffusion models • Self supervised learning • Cross-modal learning • AI • Data	**Best Match #4:** Pritam Sarkar • Deep Learning • Self-supervised Learning • Computer Vision • Multimodal • Time-series
	Best Match #5: Mohamed Harmanani • deep learning • computer vision • medical imaging • uncertainty quantification	**Best Match #5:** Mojtaba Kolahdouzi • Deep learning • Face Recognition • Image Processing • Fair machine learning

Input Student	BERT results	Groundtruth
Name: Karine Poitras **Research Interests:** • Psychology and Law • Child welfare • Child custody • Parent-child relationship Adv MRR (out of 1.0) = 0.34	**Best Match #1:** Vanessa Martin • Emotion Socialization • Adolescent Development • Parent-Child Relationships	**Best Match #1:** Nick Bala • Family Law • Juvenile Law • youth justice • child witnesses
	Best Match #2: Ian Janssen • physical activity • sedentary behavior • Children • Epidemiology • public health	**Best Match #2:** Robert Leckey • family law • constitutional law • comparative law
	Best Match #3: Nick Bala • Family Law • Juvenile Law • youth justice • child witnesses	**Best Match #3:** Vanessa Martin • Emotion Socialization • Adolescent Development • Parent-Child Relationships
	Best Match #4: Stephanie Boyer • Psychology • Medicine • Gynecology • Sexuality • Women's Health	-
	Best Match #5: Stephan Dobri • Motor function • cognitive function • Proprioception • child development • computational modeling	-

Input Student	BERT results	Groundtruth
Name: Paul Iyare **Research Interests:** • Microplastics • Plastic Pollution • Aquatic Ecotoxicology • Environmental Monitoring Adv MRR (out of 1.0) = 0.25	**Best Match #1:** Dalila CM Seckar • Limnology • Entomology • Biodiversity • aquatic toxicology • policy	**Best Match #1:** Diane M. Orihel • Aquatic Ecotoxicology
	Best Match #2: Xiaying Xin • Water and wastewater treatment • Toxicology • Emerging contaminants • Toxic bacteria • Nanotechnology	**Best Match #2:** Dalila CM Seckar • Limnology • Entomology • Biodiversity • aquatic toxicology • policy
	Best Match #3: Ahamed Ashiq • Surface Chemistry • Water Treatment • Emerging Contaminants • Waste to Energy	-
	Best Match #4: Nishana Ramsawak • water distribution systems • Biofilms • trace inorganic contaminants • water quality	-
	Best Match #5: Mohsen Zafari • Geosynthetics • Landfill engineering • Waste management • Geoenvironmental engineering	-