Citation Evidence Report

EB-1B Petition — Outstanding Professor or Researcher

8 CFR § 204.5(i)(3) · Authorship + Original Contributions

Shadi Saleh, PhD

Applied Data Scientist at Microsoft

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to the 8 CFR § 204.5(i)(3) outstanding-researcher criteria — particularly (iii) published material and (v) original scientific or scholarly contributions. It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

326	352	61	8
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

92.6% independent of 94 classified citing papers

Citation type	Count
Independent	87
Self-citation	0
Co-author	7
Same-institution	0

232 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher advanced cross-lingual information retrieval by developing methods to rerank machine-translated queries and optimizing term selection for medical domain applications.

The researcher established a foundational contribution in cross-lingual information retrieval through the 2016 paper on reranking hypotheses of machine-translated queries. This core work addresses the challenge of improving retrieval accuracy when queries are translated between languages, a critical bottleneck in global information access.

This line of work appears to address the limitations of standard translation techniques by refining how translated queries are processed and expanded. The subsequent 2019 and 2020 papers suggest an evolution toward specialized applications, specifically targeting the medical domain. By focusing on document versus query translation and term selection for expansion, the researcher extended the initial framework to handle the nuanced terminology and high-stakes accuracy requirements of medical information retrieval.

The significance of this contribution is evidenced by sustained scholarly attention. The core paper has accumulated 16 citations, while the follow-up studies have garnered 22 and 37 citations respectively. Notably, 92.6% of the 94 classified citations originate from independent researchers, indicating that this work has been widely adopted and built upon by the broader academic community rather than just the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 7 · **1 flagged influential by Semantic Scholar**

CORE PAPER

Reranking hypotheses of machine-translated queries for cross-lingual information retrieval

2016 · International Conference of the Cross-Language Evaluation Forum for European ..., 2016 · 16 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Improved cross-lingual question retrieval for community question answering	Technische Universität Darmstadt	Germany	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

FOLLOW-UP WORK

Document Translation vs. Query Translation for Cross-Lingual Information Retrieval in the Medical Domain

2020 · Proceedings of the 58th Annual Meeting of the Association for Computational ..., 2020 · 37 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Improving cross-lingual information retrieval on low-resource languages via optimal transport distillation	University of Massachusetts Amherst	United States	—
2	Soft prompt decoding for multilingual dense retrieval	University of Massachusetts Amherst	United States	**Influential**
3	Cross-lingual information retrieval from multilingual construction documents using pretrained language models	Seoul National Univ., Seoul National University	South Korea	—

No.	Citing paper	Citing institution(s)	Country	S2
4	Mixed attention transformer for leveraging word-level knowledge to neural cross-lingual information retrieval	University of Massachusetts Amherst	United States	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

FOLLOW-UP WORK

Term selection for query expansion in medical cross-lingual information retrieval

2019 · European Conference on Information Retrieval, 507-522, 2019 · 22 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Effective and practical neural ranking	University of Glasgow	—	—
2	Health Information Retrieval	Fraunhofer Institute	Germany	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

Contribution 2

Claim — Contribution 2

The researcher developed and evaluated NLP systems for health information retrieval, establishing a sustained track record in CLEF eHealth and MLIA tasks from 2014 to 2020.

The researcher's contribution centers on the development of natural language processing systems for health information retrieval, anchored by the 2014 paper 'CUNI at the ShARe/CLEF eHealth Evaluation Lab 2014.' This core work established a methodological foundation for evaluating consumer health search technologies within standardized international evaluation campaigns.

This line of work appears to address the challenge of adapting information retrieval techniques to specialized medical domains. The chronology suggests a progression from general eHealth search tasks in 2014 to specific consumer health search evaluations in 2018, and finally to machine learning approaches for COVID-19 literature in 2020. The titles indicate a consistent focus on refining system performance across evolving health information needs.

The significance of this work is evidenced by its uptake in the research community. The core 2014 paper has received 11 citations, while the 2018 follow-up has garnered 8 citations. Notably, 92.6% of the scholar's total citations originate from independent researchers, suggesting that these contributions have influenced peers outside the immediate institutional circle and co-author network.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 2

CORE PAPER

CUNI at the ShARe/CLEF eHealth Evaluation Lab 2014.

2014 · CLEF (Working Notes), 226-235, 2014 · 11 citations (GS)

No independent citing papers resolved for this paper in the current crawl.

FOLLOW-UP WORK

Cuni-mtir at covid-19 mlia@ eval task 2

2020 · MLIA COVID-19, 2020 · 2 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Multistage BiCross encoder for multilingual access to COVID-19 health information	University of Sheffield	United Kingdom	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

FOLLOW-UP WORK

CUNI team: CLEF eHealth Consumer Health Search Task 2018.

2018 · CLEF (Working Notes), 2018 · 8 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Health Information Retrieval	Fraunhofer Institute	Germany	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

Contribution 3

Claim — Contribution 3

The researcher established a standardized framework for universal syntactic dependencies, creating a widely adopted resource that has significantly influenced computational linguistics research.

The researcher's contribution centers on the development of Universal Dependencies 2.2, a core paper published in 2018. This work appears to represent a significant effort to standardize syntactic annotation schemes across diverse languages, providing a unified framework for dependency parsing. The titles suggest a focus on creating consistent, cross-lingual resources that facilitate broader comparative studies in natural language processing.

This line of work addresses the challenge of linguistic heterogeneity in computational models. By proposing a universal standard, the researcher likely aimed to overcome fragmentation in annotation practices, enabling more robust and generalizable parsing systems. The absence of follow-up papers by the same researcher indicates that this specific contribution stands as a foundational reference rather than part of an ongoing iterative series by the author.

The significance of this work is evidenced by its substantial citation count of 275. Furthermore, citation analysis reveals that 92.6% of citing papers originate from independent researchers, suggesting broad adoption across the global academic community. This high degree of independent uptake indicates that the framework has become a standard tool or benchmark for other scholars in the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 67

CORE PAPER

Universal Dependencies 2.2

2018 · 275 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Adversarial Attacks on Deep-learning Models in Natural Language Processing	Macquarie University, The University of Adelaide, Wuhan University	Australia, China	—
2	Graph Neural Networks for Natural Language Processing: A Survey	Central China Normal University, JD.COM, Nanjing University	Canada, China, United States	—
3	Deep Learning and Linguistic Representation	—	—	—
4	A primer on pretrained multilingual language models	Advanced Micro Devices Inc, IIT Madras, Indian Institute of Technology Madras	India, United States	—
5	Are all languages created equal in multilingual BERT?	Bloomberg L.P., Johns Hopkins University	—	—
6	Large-scale evidence of dependency length minimization in 37 languages	Massachusetts Institute of Technology	United States	—
7	Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing	Massachusetts Institute of Technology, University of California, Irvine	United States	—
8	JW300: A wide-coverage parallel corpus for low-resource languages	Unity Technologies	—	—
9	Many languages, one parser	Holistic Intelligence for Global Good, Oracle, University of Washington	United States	—
10	Universals of word order reflect optimization of grammars for efficient communication	Saarland University, Stanford University	Germany, United States	—
11	English intermediate-task training improves zero-shot cross-lingual transfer too	Amazon, New York University, Vector Institute	Canada, United States	—
12	Specializing word embeddings (for parsing) by information bottleneck	Johns Hopkins University, Stanford University	United States	—
13	Viable dependency parsing as sequence labeling	Universidade da Coruña	Spain	—
14	Deep contextualized self-training for low resource dependency parsing	Technion, IIT	—	—
15	Subword pooling makes a difference	Budapest University of Technology and Economics, SZTAKI Institute of Computer Science	Hungary	—
16	Complexity trade-offs and equi-complexity in natural languages: a meta-analysis	University of Tübingen, University of Zürich	Germany, Switzerland	—
17	Crosslinguistic corpus studies in linguistic typology	University of Bamberg, University of Zurich	Germany, Switzerland	—
18	LINSPECTOR: Multilingual probing tasks for word representations	Amazon, TU Darmstadt	Germany, United States	—
19	Parsing as pretraining	Universidade da Coruña	Spain	—
20	On the relationships between the grammatical genders of inanimate nouns and their co-occurring adjectives and verbs	ETH Zürich, Facebook AI Research, Johns Hopkins University	Switzerland, United States	—
21	Joint transition-based models for morpho-syntactic parsing: Parsing strategies for MRLs and a case study from modern Hebrew	Open University	Israel	—

No.	Citing paper	Citing institution(s)	Country	S2
22	Deep learning for natural language parsing	Durham University	United Kingdom	—
23	Recursive non-autoregressive graph-to-graph transformer for dependency parsing with iterative refinement	Idiap Research Institute	Switzerland	—
24	Second-order neural dependency parsing with message passing and end-to-end training	ShanghaiTech University	China	—
25	Anti dependency distance minimization in short sequences. A graph theoretic approach	Universidade da Coruña	Spain	—
26	Zero-shot dependency parsing with pre-trained multilingual sentence representations	University of Groningen	Netherlands	—
27	A simple joint model for improved contextual neural lemmatization	Allen Institute for Artificial Intelligence, Bloomberg L.P., ETH Zurich	Switzerland, United States	—
28	Klcpos3-a language similarity measure for delexicalized parser transfer	Charles University	Czech Republic	—
29	What do complexity measures measure? Correlating and validating corpus-based measures of morphological complexity	University of North Texas, University of Tübingen	Germany, United States	—
30	Cross-lingual syntactic transfer with limited resources	Columbia University	United States	—

Showing the 30 most-cited of 67 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Universidade da Coruña	Spain	SCImago #3314 · THE 1201–1500 · QS 1001-1200	10
Johns Hopkins University	United States	SCImago #33 · THE 16 · QS 24	8
Charles University	Czech Republic	SCImago #797 · THE 401–500 · QS =265	8
TU Wien	Austria	SCImago #1661 · THE 301–350 · QS =197	4
University of Massachusetts Amherst	United States	SCImago #788 · QS =247	4
SZTAKI Institute of Computer Science	Hungary	—	4
University of Washington	United States	SCImago #45 · THE 25 · QS 81	3
Uppsala University	Sweden	SCImago #349 · THE 128 · QS 93	3
Bloomberg L.P.	—	—	3
Holistic Intelligence for Global Good	—	—	2
New York University	United States	SCImago #116 · THE =31 · QS 55	2

Institution	Country	World ranking	Citing papers
University of Tübingen	Germany	THE =98	2
University of Zurich	Switzerland	SCImago #313 · QS 100	2
Oracle	Switzerland	—	2
Unity Technologies	—	—	2

Geographic distribution of citing authors

Country	Citing papers
United States	26
Germany	12
Spain	10
Switzerland	8
Czech Republic	8
Hungary	6
France	6
Austria	4
Sweden	4
United Kingdom	4
China	4
Israel	3

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- ☐ Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).
- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition's merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition's exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	Reranking hypotheses of machine-translated queries for cross-lingual information retrieval	7	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 2	CUNI at the ShARe/CLEF eHealth Evaluation Lab 2014.	2	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 3	Universal Dependencies 2.2	67	8 CFR 204.5(i)(3) — Outstanding Researcher