

7a.003.UL - Ontology-based Fast Semantic Indexing for Structured and Unstructured Data in Health Care

Project - Team

Team Member	Role	Email	Phone Number	Academic Sites/Industry Members
Satya Katragadda	PI	satya@louisiana.edu	Not Available	UL Lafayette
Raju Gottumukkala	Co-PI	raju@louisiana.edu	(337) 482-0632	UL Lafayette
Vijay Raghavan	Co-PI	vvr3254@louisiana.edu	(337) 482-6603	UL Lafayette
Adeola Olaleye Siwoku	Student	c00301223@louisiana.edu	(337) 735-0219 - Cell	UL Lafayette
Mike Lucito	Project Mentor	Michael_Lucito@Schumacherclinical.com	Not Available	Funded By: Schumacher Clinical Partners
Chris Cotteleer	Project Mentor	Not Available	Not Available	Funded By: Schumacher Clinical Partners

Attention Project PIs

PIs are responsible for keeping up the content of their project page and have the ability to EDIT the page.

- To **EDIT**, click the edit "pencil icon" in the top right-hand corner of this page
- To **PUBLISH** your changes, click the blue "Publish" button in the lower right-hand corner of this page
- If you need help or have questions, please contact Site Admin: Sally.Johnson@louisiana.edu

Project - Summary

In the current big data environment, most of the data is gathered from multiple sources. Entity resolution or duplication of data is a major problem in this scenario. This duplicate data is more pronounced in-patient data from health care. Recent studies indicate that about 15% of the Master Patient Index of major hospitals are duplicate entries. Issues like heterogeneous data, incomplete information, constantly changing properties associated with entities, and temporal information pose major challenges to identifying duplicate entities in the data. To solve this problem, we propose an indexing technique that identifies duplicate information from databases using ontology based semantic measures. The proposed approach generates a global identifier for each entity based on the distances of the properties associated with the entity to core nodes within the semantic graph extracted from the ontology. Partial and complete match algorithms will be applied on the global identifier to identify duplicate records. The identifier can be updated based on changes to the properties associated with the entity. Our project proposes a proof of concept to identify duplicate records in a Master Patient Index that indexes the data using a global patient identifier that is based on the demographic and clinical profile of the patient. We aim to significantly improve the performance of the deduplication algorithm over the traditional baseline algorithms.

Project - Novelty of Approach

1. Ontology-based approach to
 - a. Understand schema and connect objects for faster indexing
 - b. Convert, via connected objects, unstructured to structured data (e.g. records without SSN)
2. Dynamic weighted approach for calculating match probability

Project - Deliverables

Deliverable
1 Investigate various instance matching based on entity recognition, record linkage, and entity co-reference approaches in current literature
2 Develop a global identifier for each instance based on the properties or features associated with that instance
3 Design a blocking technique that identifies the matching between two instances based on the global identifier

Table of Contents

- Project - Team
- Project - Summary
- Project - Novelty of Approach
- Project - Deliverables
- Project - Benefits to IAB
- Project - Presentation Video (Spring 2018)
- Project - Documents
- Project - Comments

Spaces

- All Spaces

	CVDI 2017 IAB Fall Meeting	 +	 
	CVDI 2018 IAB Fall Meeting	 +	 
	CVDI 2018 IAB Spring Meeting	 +	 
	CVDI 2019 IAB Fall Meeting	 +	 
	CVDI 2019 IAB Spring Meeting	 +	 
	CVDI Calendar	 +	 

4 Build a prototypical system for healthcare data to identify duplicate entries in a Master Patient Index

Project - Benefits to IAB

1. Aid to streamline vital clinical and payer information such as Patient Registration, Claims Payment, Physician Referrals, etc.
2. Makes Health Information Exchange (HIE) and establishing Health Data Network (HDN) easier
3. Applies to non-healthcare business with respect to customer data

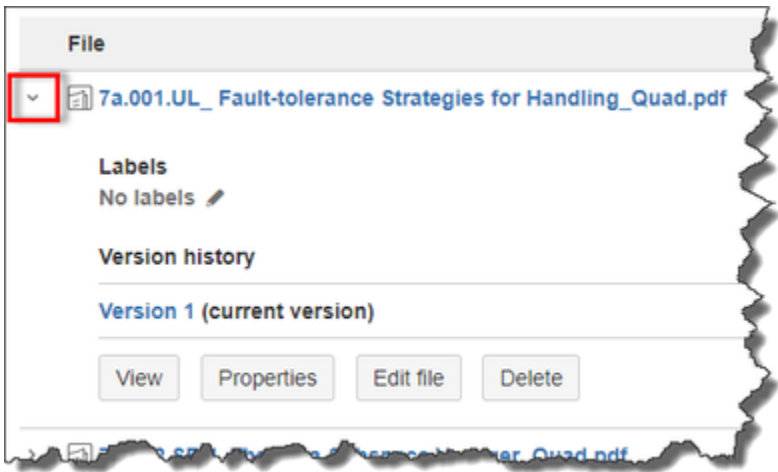
Project - Presentation Video (Spring 2018)

[Video Link \(11:46 minutes\)](#)

Project - Documents

For viewing/editing options, please click left arrow next to document name.

You will see different options depending on your access level.



File	Modified
7a.003.UL_Lightning Fast Indexing_Quad_2017 Fall Meeting.pptx	Nov 15, 2017 by Sally Johnson
7a.003.UL_CVDI UL Lafayette IAB Winter Meeting Presentation.pptx	Jan 29, 2018 by Sally Johnson
Year 7_CVDI Project Presentation 7a.003_Fast_Indexing_v3.PPTX	Feb 23, 2018 by Jian Chen
7a.003.UL_Quad Chart_2018 Spring Meeting.pptx	Mar 19, 2018 by Sally Johnson
7a.003.UL_Executive Summary_Original.docx	Jun 25, 2018 by Sally Johnson
7a.003.UL_Executive Summary_Revised after IAB Voting.docx	Jun 25, 2018 by Sally Johnson
7a.003.UL_2018 Fall Meeting Poster.pptx	Nov 14, 2018 by Sally Johnson
7a.003.UL_Mid-Year Report.docx	Jan 04, 2019 by Sally Johnson

	CVDI Leadership (All Sites)		+	★	☆
	CVDI Marketing Materials		+	★	☆
	CVDI Reports & Document Library		+	★	☆
	CVDI SITE (Drexel University)		+	★	☆
	CVDI SITE (Stony Brook University)		+	★	☆
	CVDI SITE (Tampere University)		+	★	☆
	CVDI SITE (University of Louisiana at Lafayette)		+	★	☆
	CVDI SITE (University of North Carolina at Charlotte)		+	★	☆
	CVDI SITE (University of Virginia)		+	★	☆
	IAB - Industry Advisory Board		+	★	☆
	Year 6 - Funded Projects (7/1/17 - 6/30/18)		+	★	☆
	Year 7 - Funded Projects (7/1/18 - 6/30/19)		+	★	☆
	Year 8 - Proposed Projects		+	★	☆

Drag and drop to upload or [browse for files](#)

↓ Download All

Project - Comments