



# VCU

Virginia Commonwealth University  
**VCU Scholars Compass**

---

Undergraduate Research Posters

Undergraduate Research Opportunities Program

---

2017

## Evaluating Feature Extraction Methods for Biomedical Word Sense Disambiguation

Clint A. Cuffy

*Virginia Commonwealth University*

Sam Henry

*Virginia Commonwealth University*

Bridget T. McInnes

*Virginia Commonwealth University*

Follow this and additional works at: <https://scholarscompass.vcu.edu/uresposters>

© The Author(s)

---

### Downloaded from

Cuffy, Clint A.; Henry, Sam; and McInnes, Bridget T., "Evaluating Feature Extraction Methods for Biomedical Word Sense Disambiguation" (2017). *Undergraduate Research Posters*. Poster 262.

<https://scholarscompass.vcu.edu/uresposters/262>

This Book is brought to you for free and open access by the Undergraduate Research Opportunities Program at VCU Scholars Compass. It has been accepted for inclusion in Undergraduate Research Posters by an authorized administrator of VCU Scholars Compass. For more information, please contact [libcompass@vcu.edu](mailto:libcompass@vcu.edu).

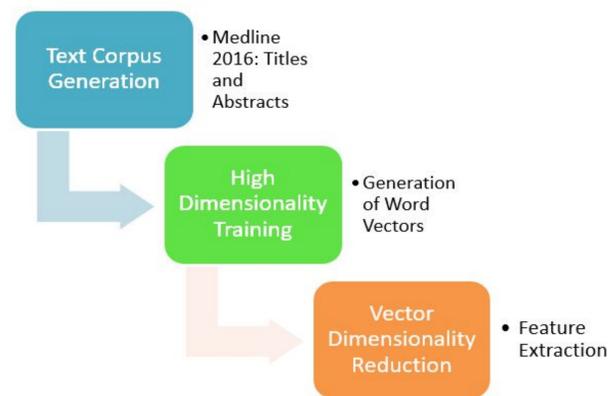


# Evaluating Feature Extraction Methods for Biomedical Word Sense Disambiguation

## Introduction

Biomedical text is a highly active research area, but ambiguity still poses a barrier to the comprehension of these documents. Many word sense disambiguation (WSD) approaches represent instances of an ambiguous word as a distributional context vector. One problem with using these vectors is noise -- information that is overly general and does not contribute to the word's representation. Feature extraction approaches attempt to compensate for sparsity and reduce noise by transforming the data from high-dimensional space to a space of fewer dimensions. In this work, we evaluate feature extraction methods for word sense disambiguation.

## Overall



## Noise Reduction

### Instance

As the heart pumps blood through the circulatory system, it flows through both atria and ventricles.

### Sense

- a hollow muscular organ that pumps the blood through the circulatory system by rhythmic contraction and dilation. In vertebrates there may be up to four chambers (as in humans), with two atria and two ventricles.
- the central or innermost part of something.

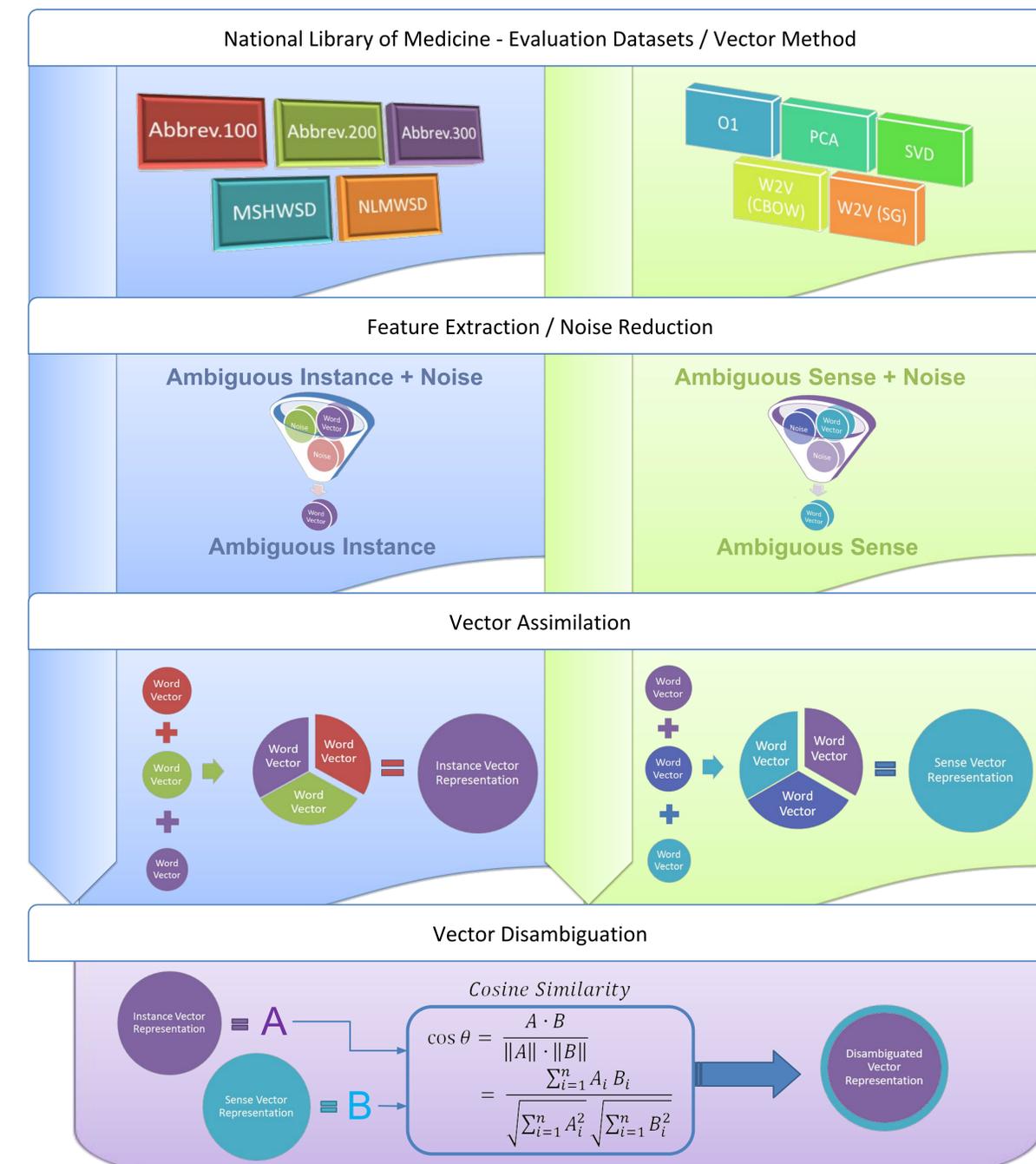
### Instance

heart pumps blood circulatory system flows both atria ventricles

### Sense

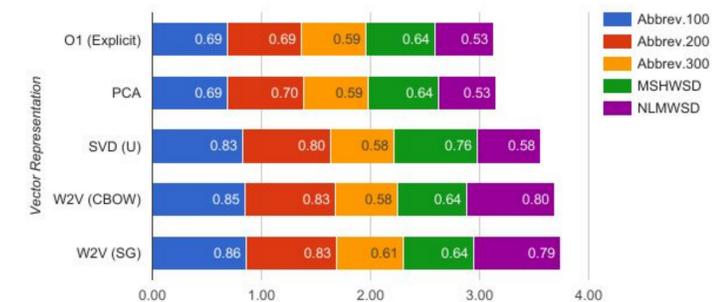
- hollow muscular organ pumps blood circulatory system rhythmic contraction dilation vertebrates may four chambers humans atria ventricles
- central innermost part something.

## Method



## Results

### Overall Accuracy Results



### Statistical Significance

	Abbrev.100				Abbrev.200			
	PCA	SVD	CBOW	SG	PCA	SVD	CBOW	SG
explicit	0.65	0.0008	0.0015	0.0006	0.29	0.006	0.0047	0.0045
PCA		0.0007	0.0013	0.0005	PCA	0.005	0.0042	0.0037
SVD			0.94	0.97	SVD		0.56	0.93
CBOW				0.93	CBOW			0.60

	Abbrev.300				MSH-WSD				
	PCA	SVD	CBOW	SG	PCA	SVD	CBOW	SG	
explicit	1.0	1.0	0.41	0.63	explicit	0.37	0.0001	0.0001	0.0001
PCA			0.41	0.63	PCA		0.0356	0.0005	0.0001
SVD			0.29	0.21	SVD			0.0005	0.0346
CBOW				0.08	CBOW				0.056

	NLM-WSD			
	PCA	SVD	CBOW	SG
explicit	0.35	0.10	0.0062	0.0127
PCA		0.087	0.0042	0.009
SVD			0.2489	0.2993
CBOW				0.66

Note: Significant Difference: Values < 0.05

## Conclusion

Although biomedical literature can be ambiguous, our results show feature extraction methods increase the accuracy in disambiguating biomedical text by eliminating variations of associated noise.

We find no statistical significance between word2vec word embeddings and SVD