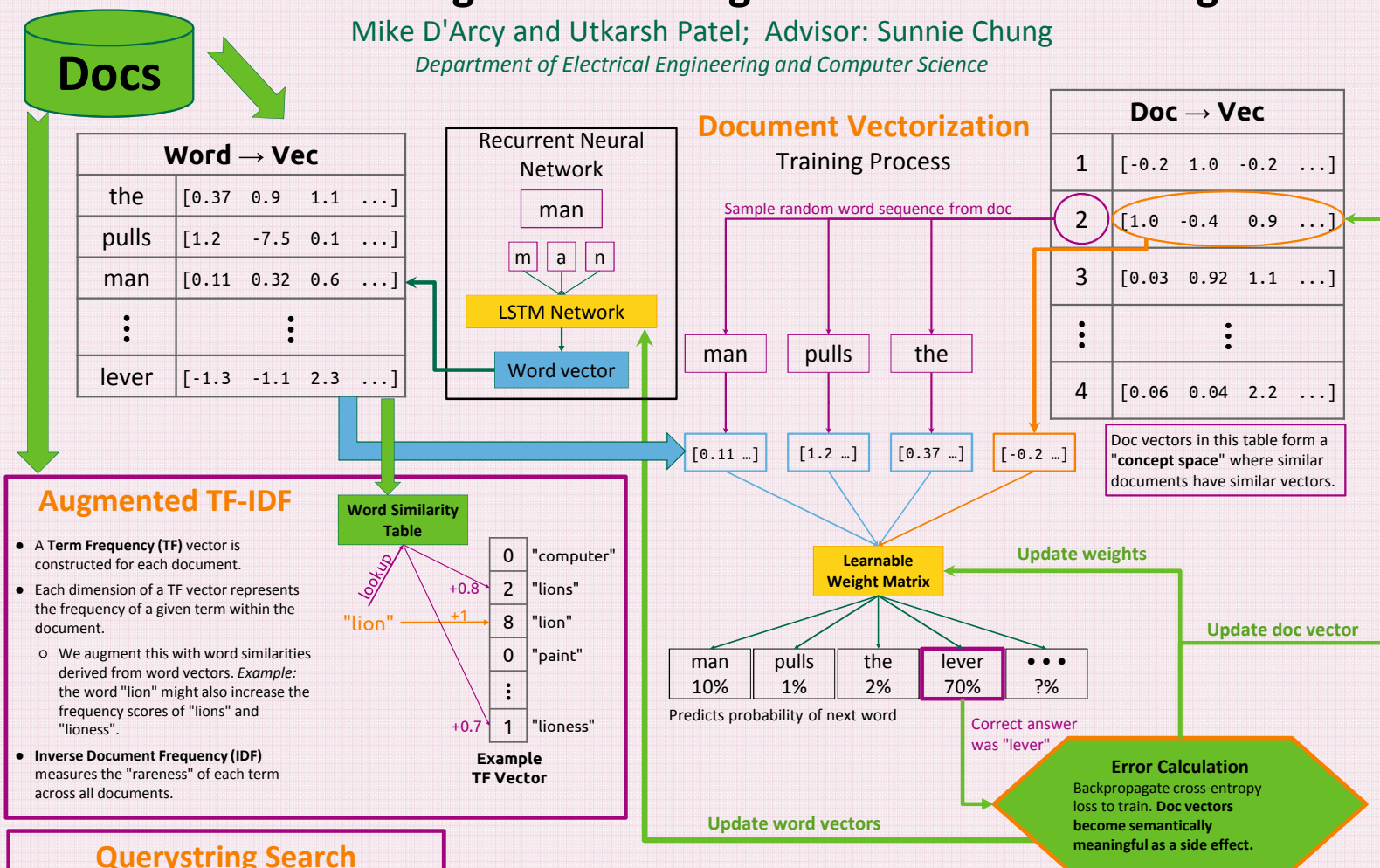


Machine Learning Based Intelligent Full Text Search Engine

Mike D'Arcy and Utkarsh Patel; Advisor: Sunnie Chung
Department of Electrical Engineering and Computer Science



Querysting Search

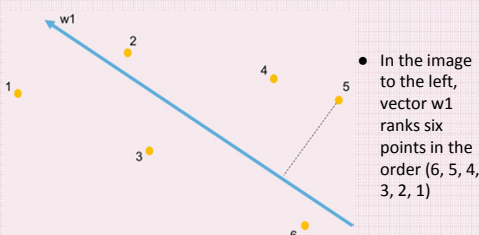
Like Google: Given a short phrase, find matching documents

- Build TF vector for query
- Compare query TF against document TFs using cosine similarity (dot product)
 - Weight by IDF (matching rarer terms matters more)

Ranking SVM

Purpose: To rank the results of query string search effectively

- Input:** A pair of documents D_i and D_k , characterized by their feature vectors ψ_i and ψ_k .
- Objective:** Learn weight vector w such that, if $\text{rank}(D_i) < \text{rank}(D_k)$, then $w(\psi_i - \psi_k) > 0$
- Intuitively, for any weight vector w , the documents are ordered by their projection onto w .

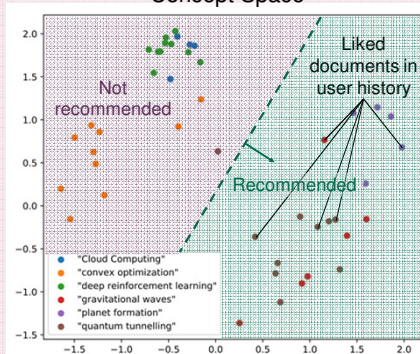


Personalized Recommendations

Based on a user's previously viewed articles, suggest new ones

- We train a **Linear Support Vector Machine** to draw a dividing hyperplane (in 2D, this is a line) through the space so all of a user's previously-liked documents lie on one side of the hyperplane.
 - Mathematically, it minimizes the **hinge loss**.
- Any unseen documents on the same side of the line are recommended to the user.

Concept Space



Related Documents

Given an article, find other articles that contain similar topics

- Document vectors already arranged spatially based on similarity
- Given a query document, just find nearest neighbor using Euclidean distance

Concept Space

