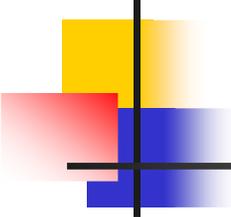


Securing Organizational Knowledge using Automated Annotation

Savitha Kadiyala

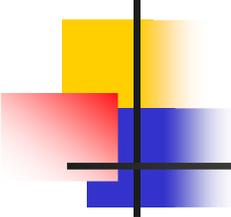
K.R. Namuduri

Venu Dasigi



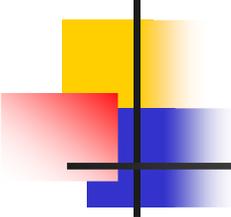
Outline

- Introduction
- Proposed model
- Relevant document retrieval
- System implementation
- Experiments
- Conclusions



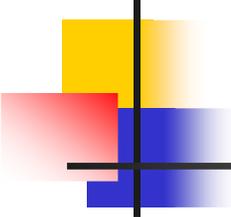
Introduction

- Defining knowledge
 - “Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information” Davenport and Prusak
- Two basic types of knowledge
 - Tacit knowledge: Uncodified knowledge deeply rooted in actions and is best described by the term “know-how.”
 - Difficult and/or expensive to transfer and hence “sticky.”
 - Explicit knowledge: Codified knowledge that can be easily communicated.
 - Easy to transfer and hence “leaky.”



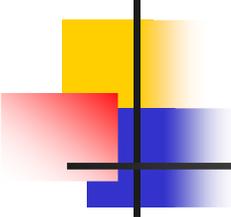
Introduction

- Safeguarding intellectual property
 - Organizational efforts concentrate on safeguarding IP from ***external*** and ***intentional*** threats such as hacker attacks or insider betrayal.
 - Patents, trademarks, copyrights, trade secrets and classifying documents.
 - The current article addresses concerns related to ***intentional*** and ***unintentional*** threats to IP that are ***internal*** to the organization.
- System checks and balances
 - Educate knowledge workers about protected IP.
 - Alert supervisors of potential violation of IP rights.



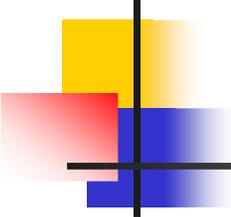
Proposed model

- Automated annotation: involves
 - Keyword extraction
 - Conducted at two places:
 - On protected documents residing in the knowledge base.
 - Knowledge workers' email drafts
 - Correlation analysis
 - Between the keywords extracted from both of these places.
 - Keyword extraction
 - Categorizing IP related documents into clusters using standard document classification techniques.
 - Vector Space Model (VSM)
 - Latent Semantic Indexing (LSI)
 - Indexing the clusters of documents using metadata.
 - Extracting keyword from the drafted email.



Fit with the seven-layered architecture for knowledge management systems

Interface layer (Browser)
Access and Authentication Layer (authentication, security, recognition)
Collaborative Intelligence and Filtering Layer (keyword extraction tools, meta tagging tools)
Application Layer (Skills directories, yellow pages, digital work rooms, video conferencing)
Transport Layer (TCP/IP, Document exchange, Email, Video transport)
Middleware and Legacy Integration Layer (tools that integrate data across platforms)
Repositories (Knowledge bases storing IP protected documents)

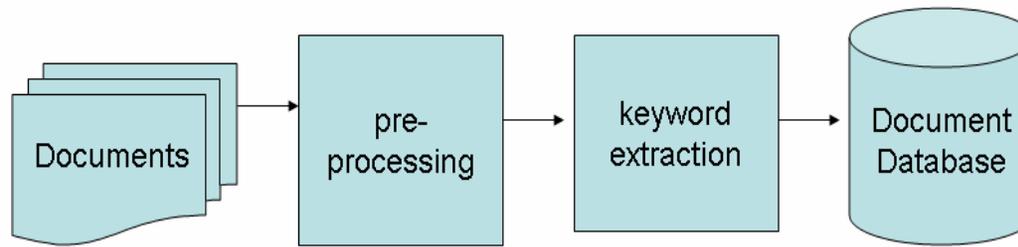
$w_{i,j}$ 

Relevant document retrieval

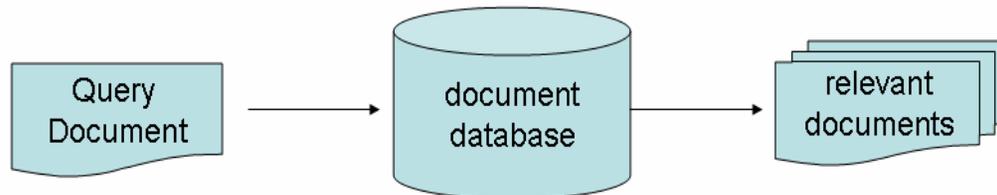
$$w_{i,j} = \frac{\text{freq}_{i,j}}{\max_l \text{freq}_{l,j}}$$

$$\begin{aligned} \text{sim}(\vec{d}_j, \vec{q}) &= \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} \\ &= \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{j=1}^t w_{i,q}^2}} \end{aligned}$$

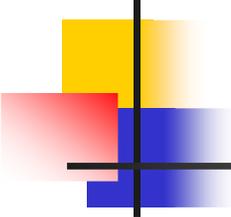
System Implementation



(a) Acquisition Phase



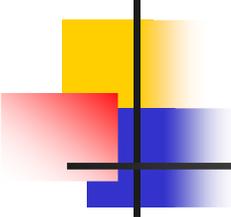
(b) Monitoring Phase



Experiments

- Compact representation of a sample patent application.

• 0.0435	• aqueous
• 0.1304	• chloride
• 0.0435	• crystals
• 0.0435	• disintegrate
• 0.0870	• dosage
• 0.0435	• ethylcellulose
• 0.1304	• granules
• 0.1304	• invention
• 0.0435	• milliequivalents
• 0.2609	• potassium
• 0.0435	• surfactants



Conclusion

- Automated annotation is used to recognize and alert knowledge workers about potential security threats to organizational knowledge.
- This model is implemented with the aid of a set of document processing tools and document search and retrieval tools.
- One of the challenges in implementing the system is to balance the knowledge workers' need for email privacy as well as the organizational need to safeguard its IP.