

# Silverfish

A web-based semantics extraction and aggregation engine for academic documents

Dec 18, 2008

# Overview

- 1 Introduction
- 2 Architecture
- 3 Features
- 4 Roadmap

# What is Silverfish

- A web-based semantics extraction and aggregation engine for academic documents
- Contribution to Open Courseware Consortium from IIT-B
- Automatically extracts entities from academic documents
- Aggregated knowledge is presented to the user through a web interface
- Silverfish can be accessed at <http://silverfish.iitb.ac.in>

## More on Silverfish

- Documents can either be submitted by the user or will be crawled on request.
- Built as a Social platform where users can interact
- Acts as a recommendation engine for relevant content.
- Intended to be a **One Stop Shop for Academicians and Researchers**

# Overview

- 1 Introduction
- 2 Architecture
- 3 Features
- 4 Roadmap

# Architecture

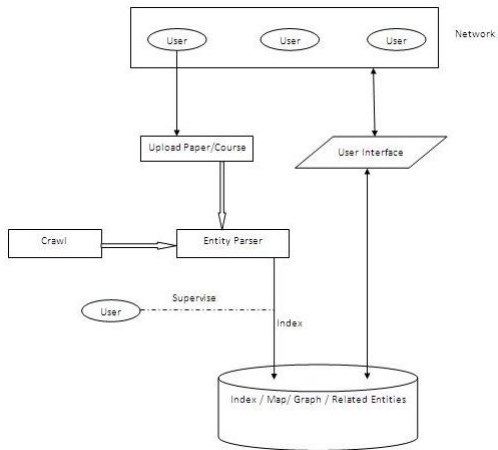


Figure: Architecture of Silverfish

# What Silverfish Extracts

- Paper Title — Course Page Title
- Authors of the Papers — Professors offering the course
- Abstract of the Paper — Description of the Course
- Institute
- Keywords

## How the Entities are stored

- Extracted entities are stored as an evolutionary co-occurrence graph with undirected weighted edges
- Users are also modeled as Entities in the graph
- All entities extracted from a paper form a clique
- Other Edge types include a user's network, favorites, equivalence edges etc



# Extracted Entities

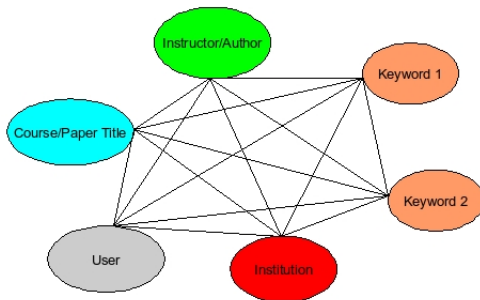


Figure: Extracted Entities Clique

# Backend Visualization

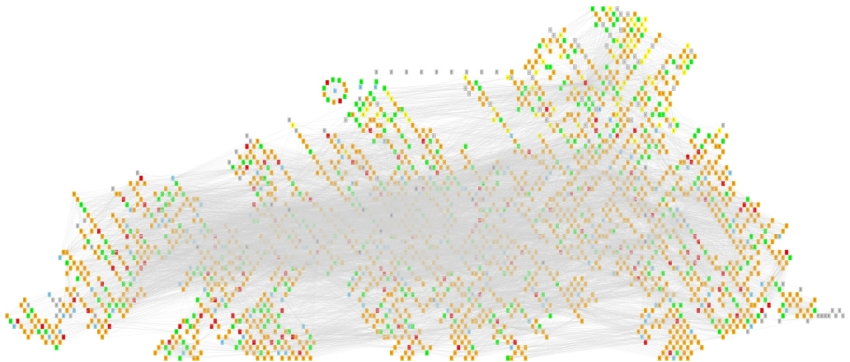


Figure: Visualization of Silverfish backend (Radius of the graph = 3)

# Overview

- 1 Introduction
- 2 Architecture
- 3 Features**
- 4 Roadmap

# Features

- Upload: User can upload papers, course pages into Silverfish.
  - Extracted Entities are presented to the user for correction/addition and are finally added to the graph
- **Users can build and maintain a personal repository of academic documents**
- Users can favorite Entities that they discover in the system.
- Social Network: User can network with other users
- Related Entities: A user viewing an entity is shown related entities.

# Features

- **Users get updates about Silverfish activity on their profiles**
  - Updates include new papers/courses uploaded in relevant fields, new network connections, annotations on uploaded papers etc
  - Updates are based on their favorites, uploaded papers and network.
- **Search: Full text search on all entities.**
  - Suggestions are provided on search terms
  - Related entities are also shown in the search results
- **CFP: Users can submit a CFP for a conference. Silverfish will try and extract topics from the CFP and send invitations to only relevant Silverfish users.**

# Overview

- 1 Introduction
- 2 Architecture
- 3 Features
- 4 Roadmap**

# The Road Ahead

- Do more with citations and references
- Improve relevance of recommendations presented to users
- Include other academic artifacts like Journals, Books etc
- Recommend User's with similar interests