

## *Amalthea*: Information Discovery and Filtering using a Multiagent Evolving Ecosystem

**Alexandros Moukas**  
**MIT Media Lab.**

발표: 정순철

## Question

The two most important features of Amalthea?

## Motivation

- Information overload
- Suited to this domain
  - ┆ Automatical evolution rate control
  - ┆ Quick adaptation to a user's new interest
  - ┆ Exploration VS Exploitation
- Various application domains
  - ┆ WWW (using search engines)
  - ┆ information streams (news server)
  - ┆ site monitoring

## Features

- Digest reporting
- Background operation
- Multiple agents
  - ┆ Information Filtering Agents (IFAs)
  - ┆ Information Discovery Agents (IDAs)

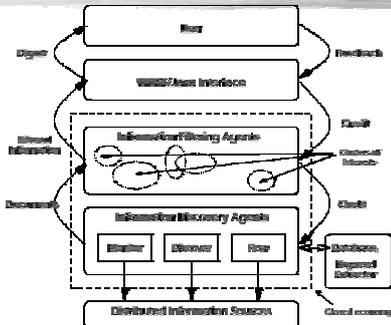
## Features

- Bottom-up approach for intelligent behavior
  - ┆ An user usually has some overlapping interests.
  - ┆ Very simple agents cooperate and compete with one another
  - ┆ Distributed adaptation
- Training agents using an genetic algorithm

## Bootstrapping Amalthea

- Generation of filtering agents
  - ┆ Bookmarks
  - ┆ Browsing history file
  - ┆ Pointing a specific page
  - ┆ Pretrained packages of agents
- Generation of discovery agents
  - ┆ Random assignment of WWW search engines

## Amalthea's Architecture



## User Interface - Configuration

The screenshot shows the 'Monitoring Amalthea's Operation' configuration window. It is divided into two main sections:
 

- Information Filtering Agents**:
  - Number of Agents: 200
  - Performance: 15
  - Current Generation: 412
  - Cloning %: 25
  - Crossover %: 5
  - Mutation %: 10
- Information Discovery Agents**:
  - Number of Agents: 25
  - Performance: 23
  - Current Generation: 412
  - Cloning %: 5
  - Crossover %: 15
  - Mutation %: 5

 At the bottom, there are buttons for 'OK', 'Defaults', and 'Cancel'.

## User Interface - Site Monitoring

The screenshot shows the 'Monitoring Sites' configuration window. It includes the following fields and options:
 

- URL to add**: A text input field containing 'http://www.meds.mt.edu' and a 'Go to the URL' button.
- Name of site**: A text input field containing 'MIT Medco Laboratory'.
- Depth to search**: A text input field containing '2'.
- Search every**: A text input field containing '2' and a 'Days' dropdown menu.
- Notification parameters**:
  - Notify when**: A text input field containing '10' and a 'percent of URL text changes' label.
  - Notify when links added/deleted to URL text**: A checked checkbox.

 At the bottom, there are buttons for 'Submit', 'Clear', and 'Cancel'.

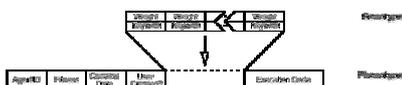
## Document Representation

- Weighted keyword vector representation
  - Preprocessing
  - Weighting keywords by TFIDF measure
  - Adjusting weights with HTML tags

$$idf_k = \log\left(\frac{N}{df_k}\right)$$

$$W_k = H_c \cdot T_f \cdot idf_k$$

## Information Filtering Agents



### General Properties

- An IFA represents a very specific interest of the user.
- Similar agents are clustered together.
- An user-creating agent is assumed to represent a long-term interest of the user. The agent is not destroyed easily.

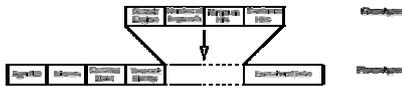
## Information Filtering Agents

### Operation

- request documents whose types they are interested in.
- filter documents using the distance between two keyword vectors a, b

$$DIF_{A,a,b} = \sqrt{\frac{\sum_j M^{ak} \cdot W^{bk}}{\sum_j (M^{ak})^2 \cdot \sum_j (W^{bk})^2}}$$

## Information Discovery Agents



- Genotype
  - Search Engine / information source to use
  - Number of keywords
  - Query type (“AND”, “OR”, etc)
  - Minimum Hits, Maximum Hits

## Interaction between IFAs and IDAs

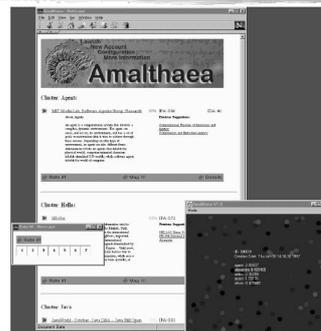
- IFAs post their requests in the request list.
- Each IDA selects the request of a suitable IFA by checking its transaction history. The IDA erases the request in the request list. (It implies n:1 relationships )
- IDAs get documents.
- IDAs presents corresponding IFAs retrieved documents.

## The Digest and Relevance Feedback

- Making the digest
  - Each IFA submits one document to Amalthea.
  - Amalthea shows the user top-ranked documents only with high confidence level.
- Relevance Feedback
  - Site rating
  - Picking keywords that best describe the given document
  - Monitoring the user’s behavior

$$C_i = D_{IFA} \cdot F_i$$

## User Interface - The Digest



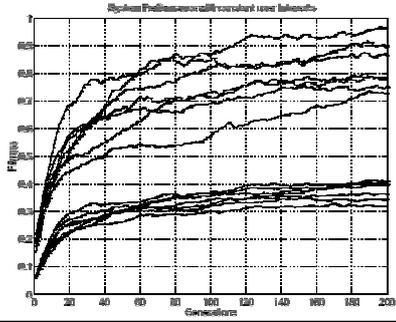
## Ecosystem

- Credit Assignment
  - IFAs receive credits based on the user’s feedback
  - IFAs give part of the credits to corresponding IDAs.
- Its credit is Its fitness.
- Evolution happens every given amount of generations.

## Evolution (Learning)

- Selection: only the fittest agents allowed
- Crossover: 2-point crossover
  - The child inherits parts of the genotypes from parents
- Mutation: the genotype of an agent + a randomly selected, randomly weighted keyword from other agents or highly-rated documents
- Cloning: just copying good agents

## Experimental Results



## Experimental Results

