

# **Web Services and Application of Multi-Agent Paradigm for DL**

Yueyu Fu & Javed Mostafa  
School of Library and Information Science  
Indiana University, Bloomington  
2005

# Outline

- Background
- Centralized vs. Distributed Classification
- Multi-agent Classification
- Discussion

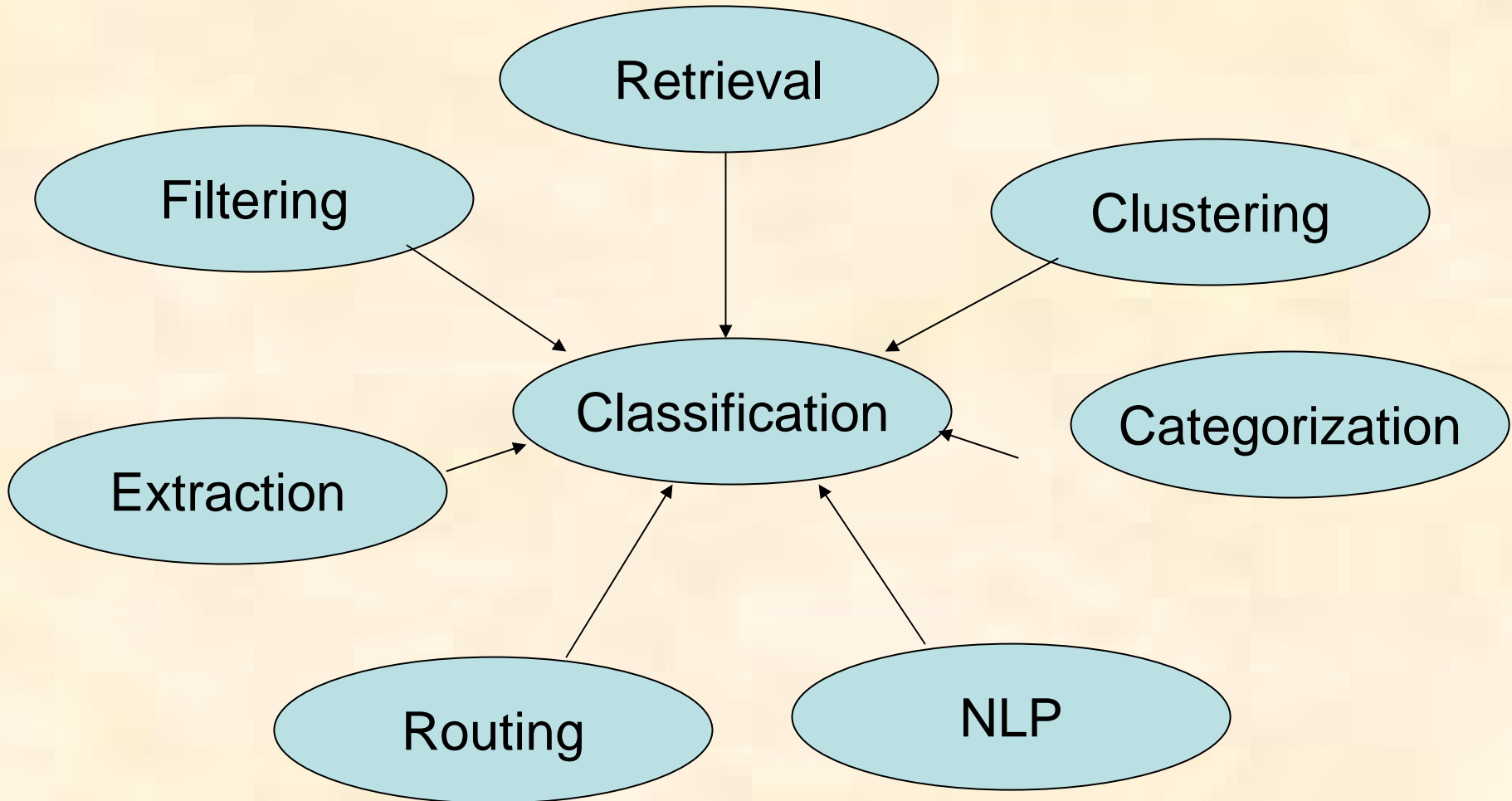
# Background

- Information overload
  - MEDLINE database contains over **12 million** records dating back to the mid-1960's.
  - Google claims that it can search more than **8 billion** web pages, which is only a small fraction of the whole web.
- Information organization
  - Document classification
- ***Document classification*** is an important operational problem in digital library research.

# What is document classification?

- Document classification --- a process of assigning natural language texts to predefined categories.
  - a news article about a basketball game – sport
  - a patent document about computer chips - technology
  - a new article about war in Iraq – politics/economic

# Why classification is important?



# Document Classification

- Human/machine
  - Manual classification
  - Automatic classification
- Organization structure
  - Centralized classification
  - Distributed classification

# Manual Classification

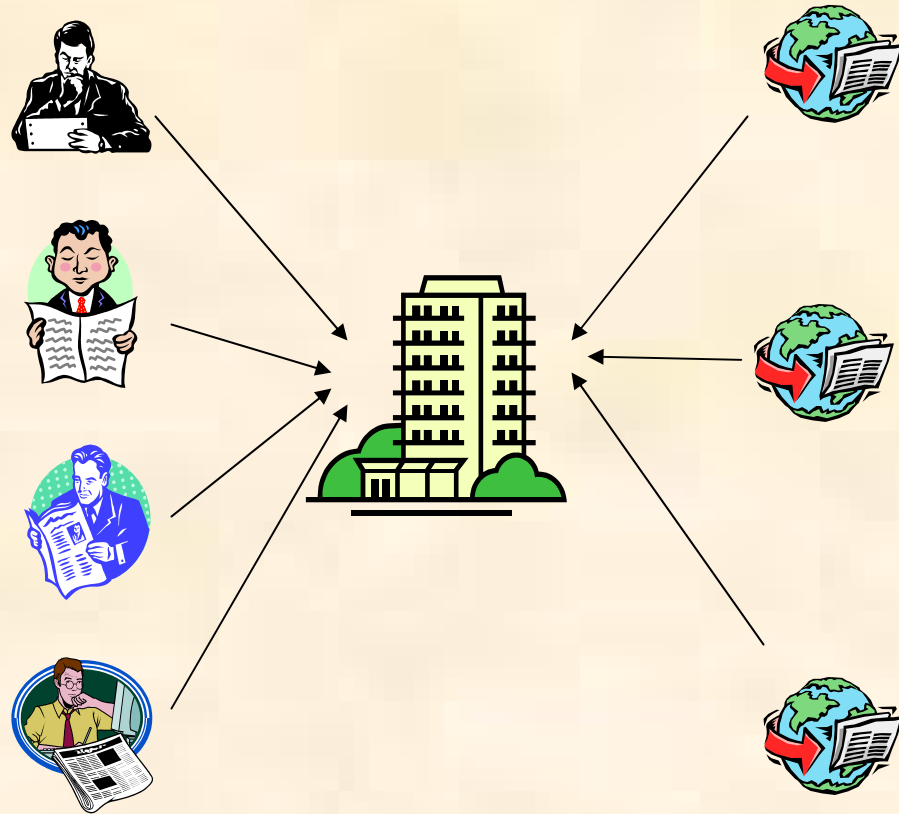
- Traditional approach – Manual classification
  - Principal schemes:
    - Dewey Decimal Classification
    - Universal Decimal Classification
    - Library of Congress Classification
  - Con: heavily rely on domain experts and human judgments

# Automatic Classification

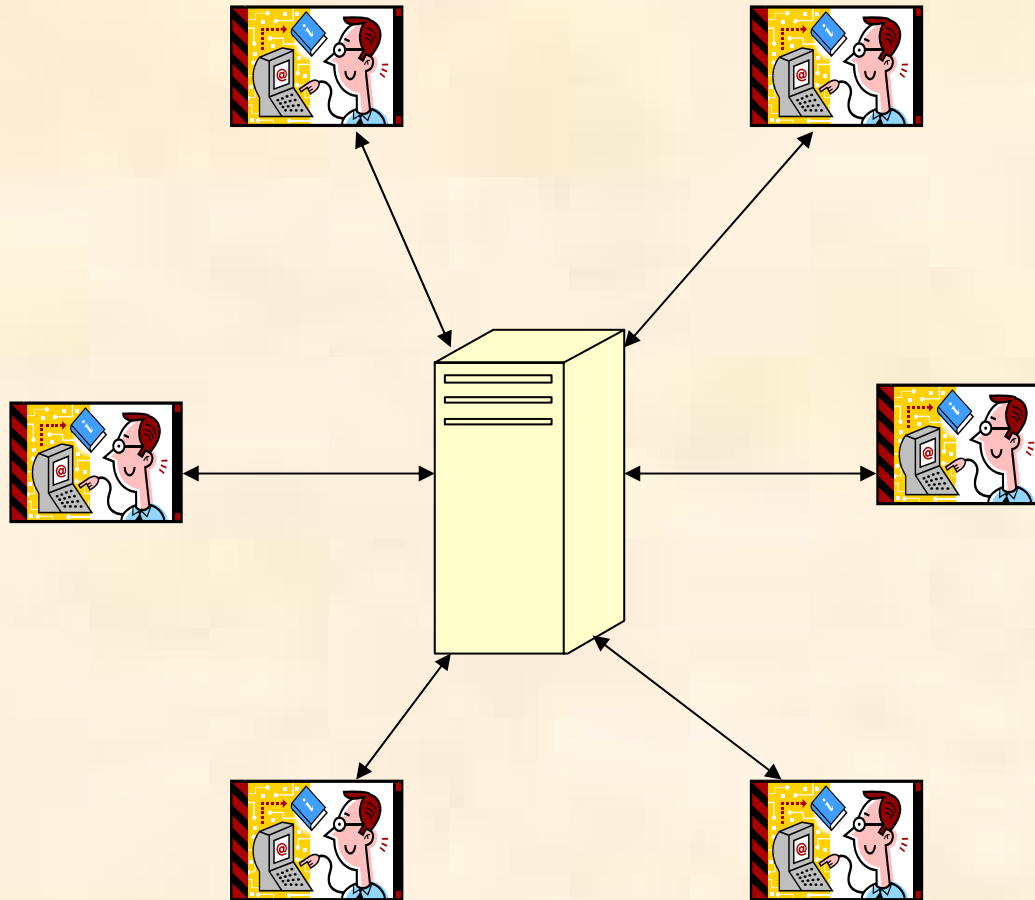
- Alternative approach – Automatic classification
  - Automatically classify texts based on a set of pre-classified documents using machine learning techniques
  - Classifiers built in a centralized and monolithic manner
  - Pro: automation, efficient, and consistent.



# Centralized Classification



# Distributed Classification



# Centralized vs. Distributed Classification

- Centralized approach
  - Classify the documents independently using a centralized and monolithic classification program
- Distributed approach
  - Allows multiple classification programs to work together to classify the documents in a distributed computing environment

# Centralized Classification: disadvantages

- Limited by its knowledge
- Limited by its computing power
- Performance bottleneck
- Single point of failure

# Distributed Classification: advantages

- More knowledge
- More computing resources
- Reliable --- avoid single point of failure
- Scalable --- dealing with large data set

# Multi-Agent Paradigm

- Evolved from distributed artificial intelligence in the late 80's
- Multi-agent system (MAS) is “a loosely coupled network of problem solvers that work together to solve problems that are beyond their individual capabilities.” (Durfee & Montgomery, 1989)

# MAS: characteristics

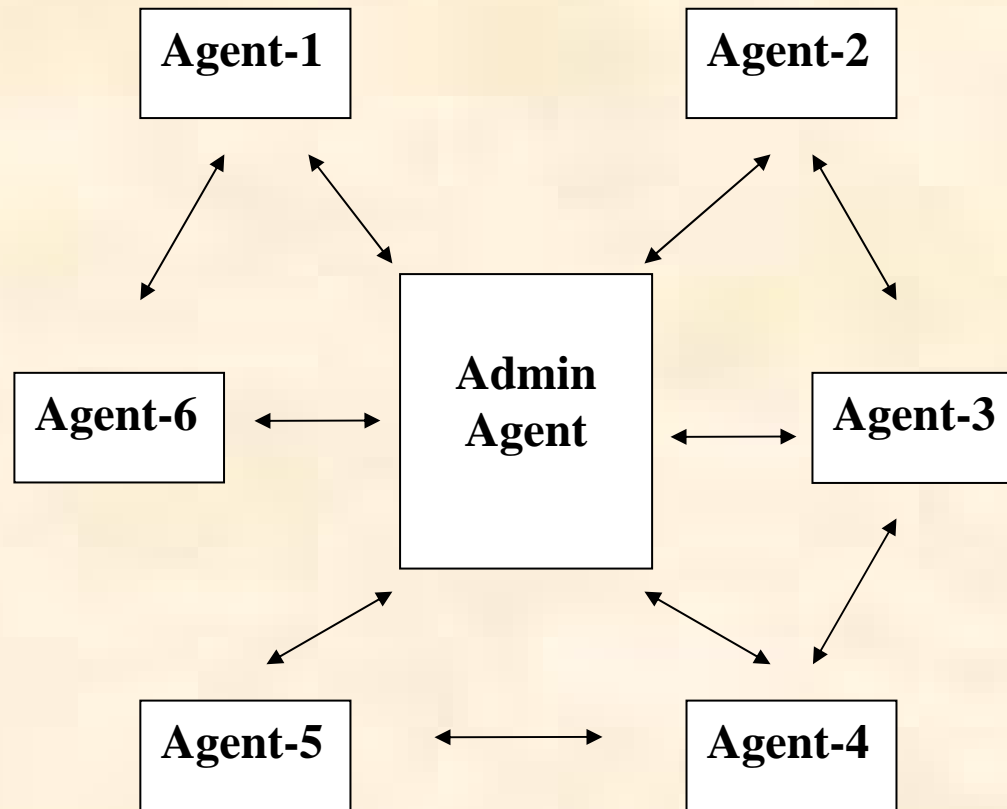
- Composed of multiple autonomous components, called agent
- Each agent has incomplete capabilities to solve a problem
- No global system control
- Data is decentralized
- Computation is asynchronous

# MAS: advantages

- Distributes computational resources and capabilities across a network of interconnected agents
- Avoids the “single point of failure” problem
- A modular, scalable architecture.
- Solutions to problems that can naturally be regarded as a society of autonomous interacting components-agents.
- Solutions that efficiently use information sources that are spatially distributed.
- Solutions in situations where expertise is distributed.
- Enhances overall system performance.



# Multi-Agent Collaboration and Classification of Information (MACCI)

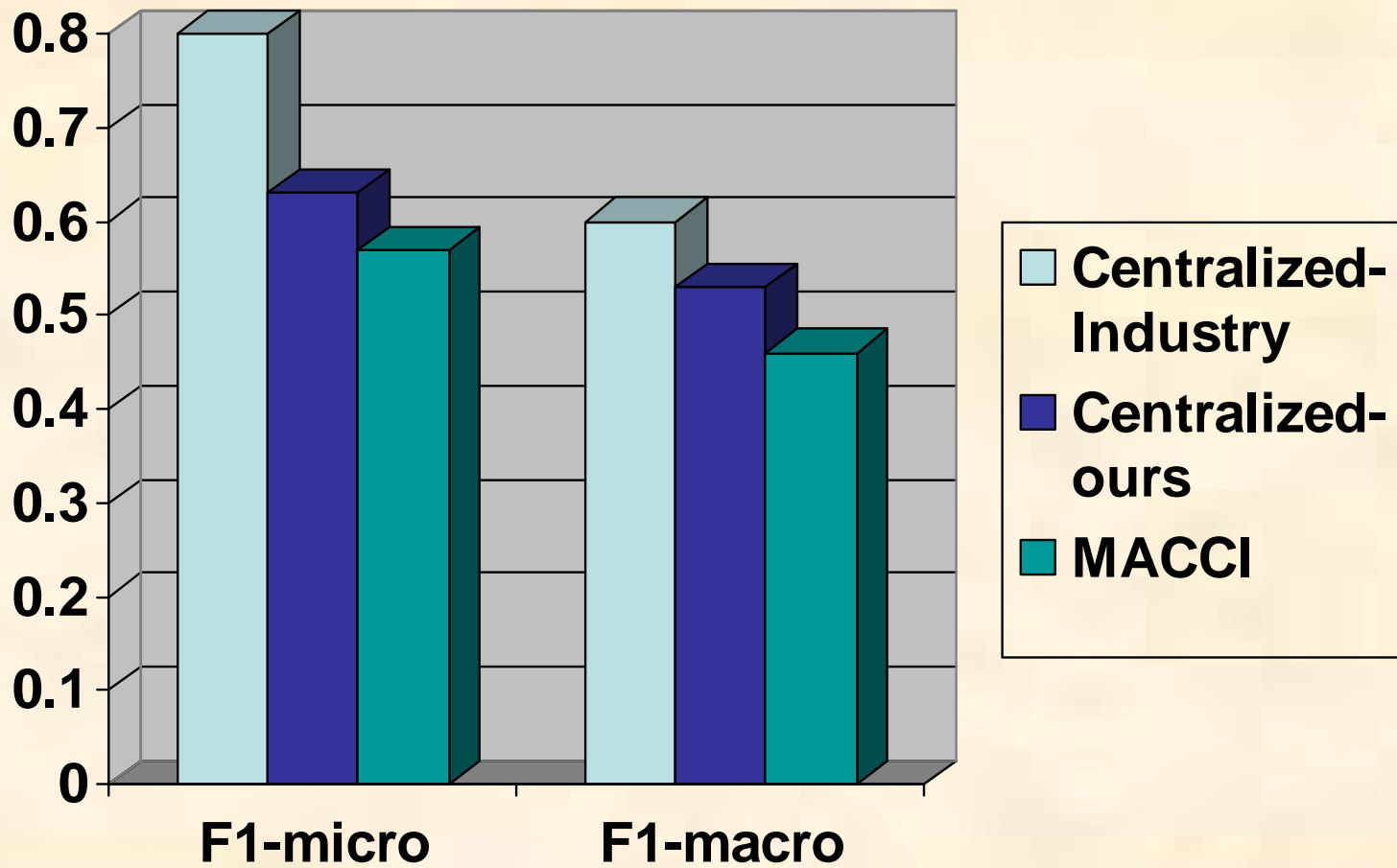


# MACCI - Experiment

- Data set
  - RCV1-v2: 800,000 manually categorized newswire stories from Reuters, Ltd.
- Classification method
  - Cosine similarity
- Effectiveness measure

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

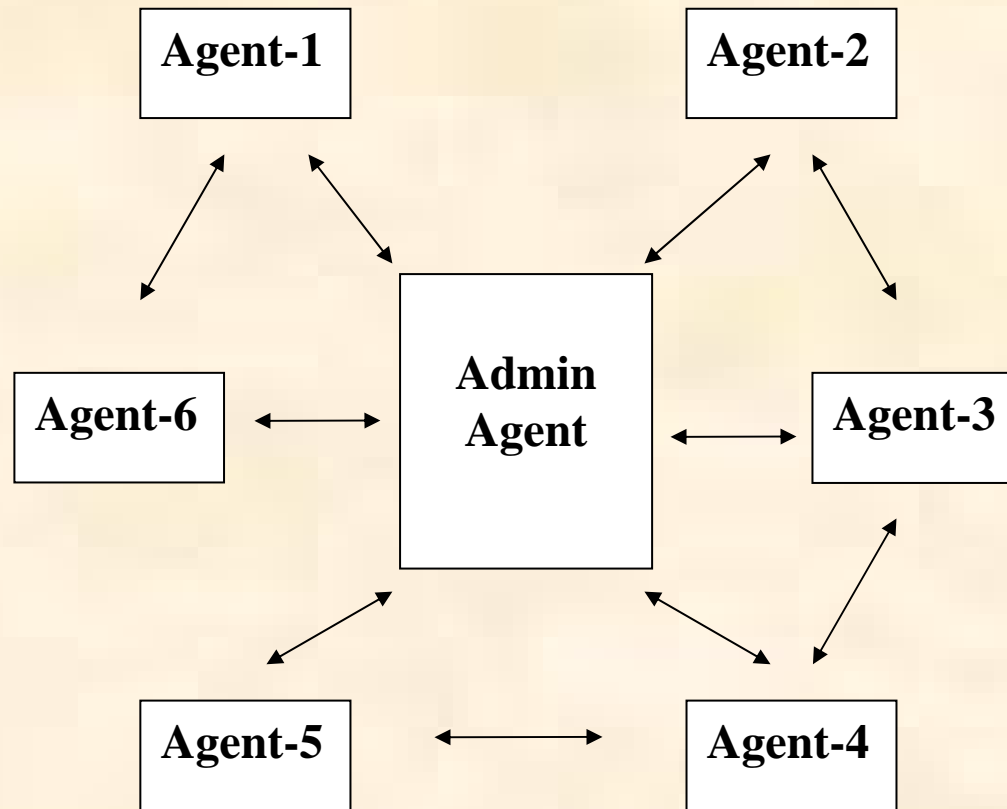
# MACCI - Results



# Agent Collaboration

- Agents collaborate to help each other.
- Agent communication and interaction are controlled by agent collaboration strategies.
- Collaboration strategies
  - Random strategy
  - Good-Neighbor strategy

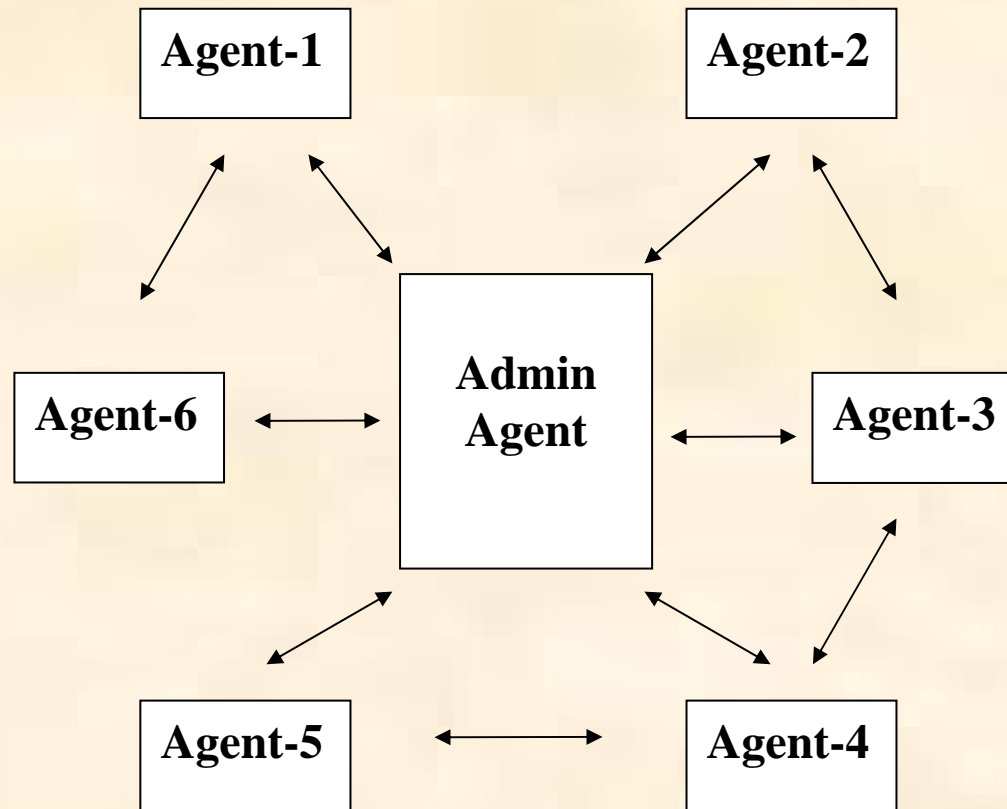
# Multi-Agent Collaboration and Classification of Information (MACCI)



# Random Strategy

1. An agent (A) asks another agent (B) for help randomly when it fails to classify a document.
2. Then agent B tries to classify the document and report the result to agent A.
  - i. If agent B classifies the document successfully, then this task is finished;
  - ii. If agent B fails to classify, agent A will repeat the steps to ask other agents for help until the document has been classified or all the other agents in the environment have been asked.

# Multi-Agent Collaboration and Classification of Information (MACCI)

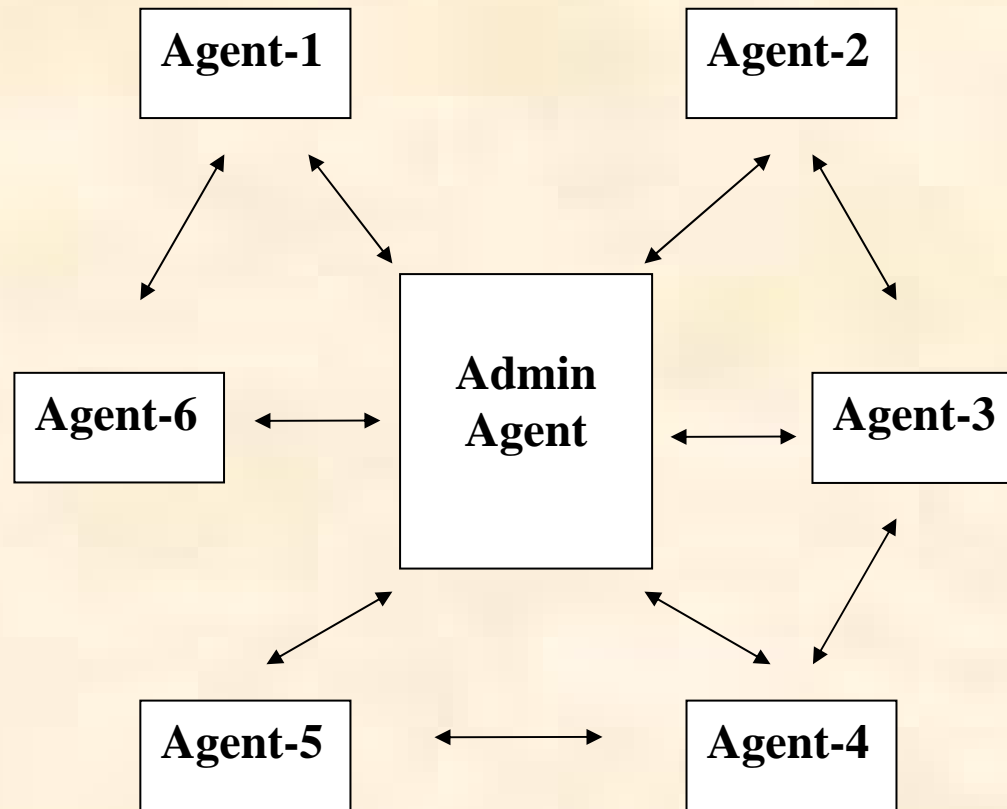


# Good-Neighbor Strategy

1. The administration agent distributed a document from the document pool to a randomly chosen classification agent. The process continues until the document pool is empty.
2. If an agent successfully classifies a document sent from the administration agent, it sends the document to all the agents in its success list for other potential classification. The help degree is set to 1.
3. If an agent fails to classify a document sent from the administration agent, it sends the document to the four top level parent agents in its failure list for help. The help degree is set to 1.
4. If an agent successfully classifies a document sent from another classification agent and the help degree is smaller than 4, it sends the document to the agents that represent its child classes in its success list. The help degree is incremented by 1.
5. If an agent fails to classify a document sent from another classification agent, it doesn't take any action.



# Multi-Agent Collaboration and Classification of Information (MACCI)



# Conclusion

- The multi-agent approach can successfully achieve the same level of effectiveness for document classification as the centralized approaches do.
- High level of effectiveness can be achieved by adapting carefully designed agent collaboration strategies.