



# **RESEARCH ON MULTIMEDIA COMMUNITY-BASED QUESTION ANSWERING**

**Name: Quan YUAN**

**School: SCE/IMI**

**Supervisor: Gao CONG, Nadia Magnenat  
THALMANN**

# CONTENT

- Motivation
- Objectives
- Problems and Methodology
- Completed Work



# COMMUNITY-BASED QUESTION ANSWERING

- Community-based Question Answering (CQA) is a popular social network that provides a platform for a large number of users to ask, answer and retrieve questions on diverse topics.
- E.g. Yahoo! Answers(YA), Naver, AnswersBag
- Limitation: unitary input modality-text.
- “A picture is worth a thousand words.”



# SCENARIO

- A lady is selecting a skirt alone in a shopping mall.
- She is not sure whether it suits her, and turns to CQA for help.

## Approach 1

I am a 22-year old girl. I want to buy a skirt for a party. I find a beautiful one but I don't know whether it suits me. It's a long black braces skirt with white followers. It can cover my calves. There is a zipper on the back. The skirt is made of silk...

I have long brown hairs and fair skin. I do not wear glasses. I am pretty slender. My style is...

## Approach 2



# OBJECTIVES

- Design a Multimedia Community-based Question Answering



# FRAMEWORK

<u>Recommendation</u>		<u>Classification &amp; Browsing</u>	<u>Retrieval (Text and Image)</u>		<u>Question Pushing</u>	
3D simulation	Sentiment Analysis	Question Organization	Storage & Indexing	Source Identification	Friend Identification	User Ranking
Feature Extraction	Opinion Question Identification				Profiling	Expertise Computation
Resolved Questions Data				Other Data Source (Flickr,		User Data

# BROWSING & CLASSIFICATION

- Questions are organized in taxonomy
- **Objective:** Developing a better taxonomy for browsing and classification
- **Methodology:**
  - Optimizing taxonomy for text-based questions
  - Creating taxonomy for multimedia questions

Which category should I post the question into?



# QUESTION RETRIEVAL

- Save time and labor
- **Objective:** Developing a multimedia question retrieval method
- **Methodology:**
  - Index
  - Identify the data resource
  - Efficient Retrieval
    - Text
    - Image
    - Bridge the two

Is there any question about selecting skirt for party? Has this skirt been asked before?

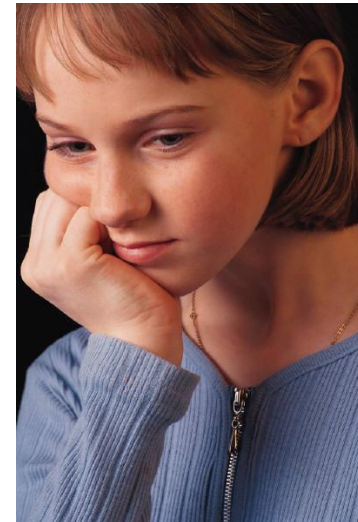




# PUSHING QUESTION

- **Objective:** Developing an approach to finding experts and pushing questions.
- **Methodology:**
  - Computing users' expertise
  - Profiling friends' interests
  - Ranking
  - Pushing the question
    - To users on CQA
    - To friends on other social network like Facebook

The party will start in 1 hour. I can't wait...



# RECOMMENDATION VIRTUAL TRYING ON

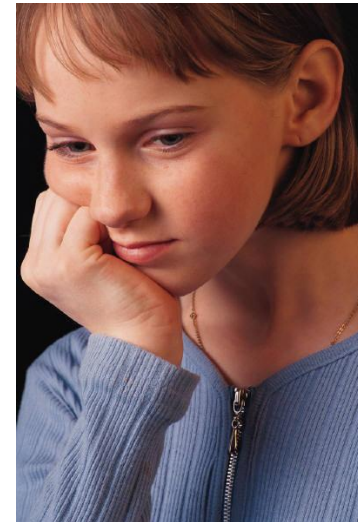
## ○ Objective:

- Develop an approach to recommendation and virtually trying on

## ○ Methodology:

- Extract features of people and clothes
- Identify the opinion questions
- Analyze the sentiments of answers
- 3D simulation

Is there any better choice?  
How about the skirt on that girl?



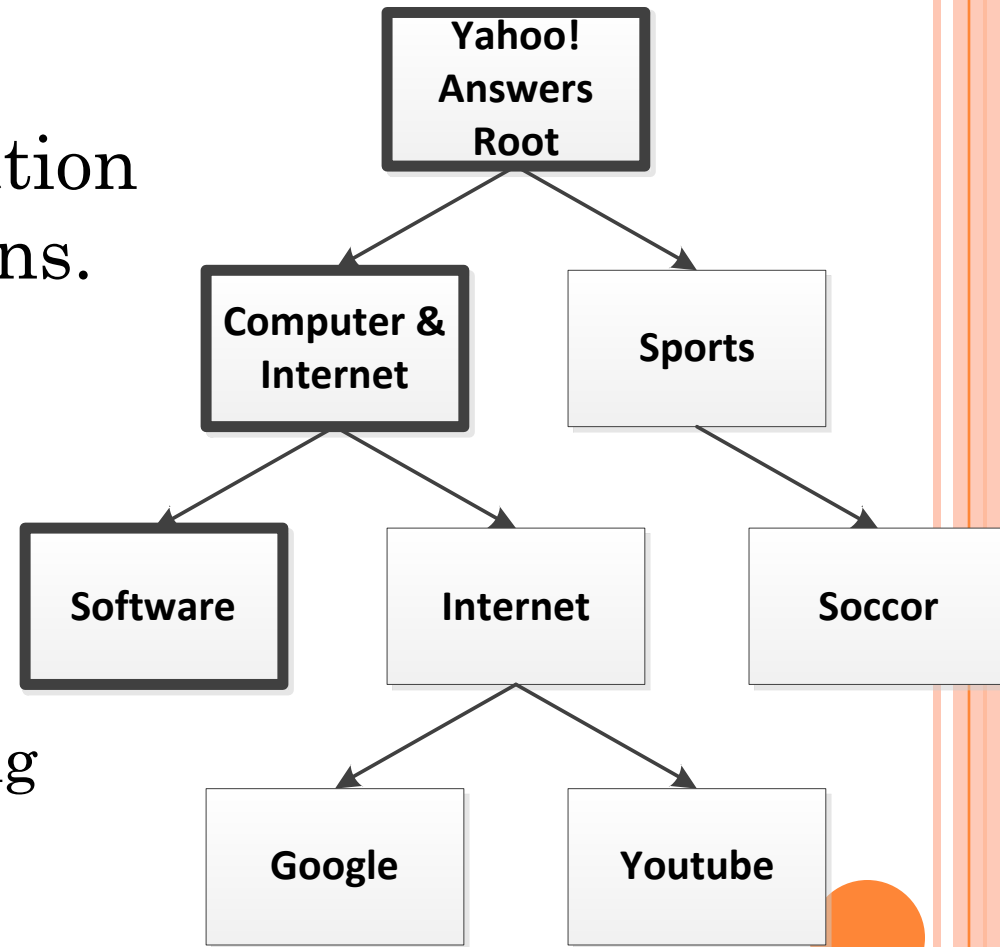
# COMPLETED WORK

- Learned basic techniques used in CQA.
- Learned some methods of image retrieval.
- Finished a paper– “An automatic approach to optimizing the taxonomy of CQA”



# THE TAXONOMY OF CQA AND ITS FUNCTIONS

- Taxonomy is a hierarchical organization of documents/questions.
- Taxonomy is very useful
  - Facilitates browsing;
  - Facilitates categorizing



A small part of taxonomy of Yahoo! Answers (YA)

# PROBLEMS OF TAXONOMY OF CQA

## ○ Problems:

### ○ Poor Granularity:

- Coarseness: not specific enough;
- Over-fine: too detailed.

### ○ Overlap: Not distinct enough

## ○ Results:

- Poor classification performance.
- Hard to use.

## ○ Our Goal:

Given a taxonomy  $T$ , we aim to find a new taxonomy  $T_n$ , such that the classification performance of  $T_n$  is better than that of  $T$ .



# RELATED WORK

- <On the merits of building categorization systems by supervised clustering>
- <Automatically learning document taxonomies for hierarchical classification>
- <A practical web-based approach to generating topic hierarchy for text segments>
- <Acclimatizing Taxonomic Semantics for Hierarchical Content Classification>

**Weak points:** all failed to use existing semantics taxonomy, use external resource and re-construct leaf nodes.

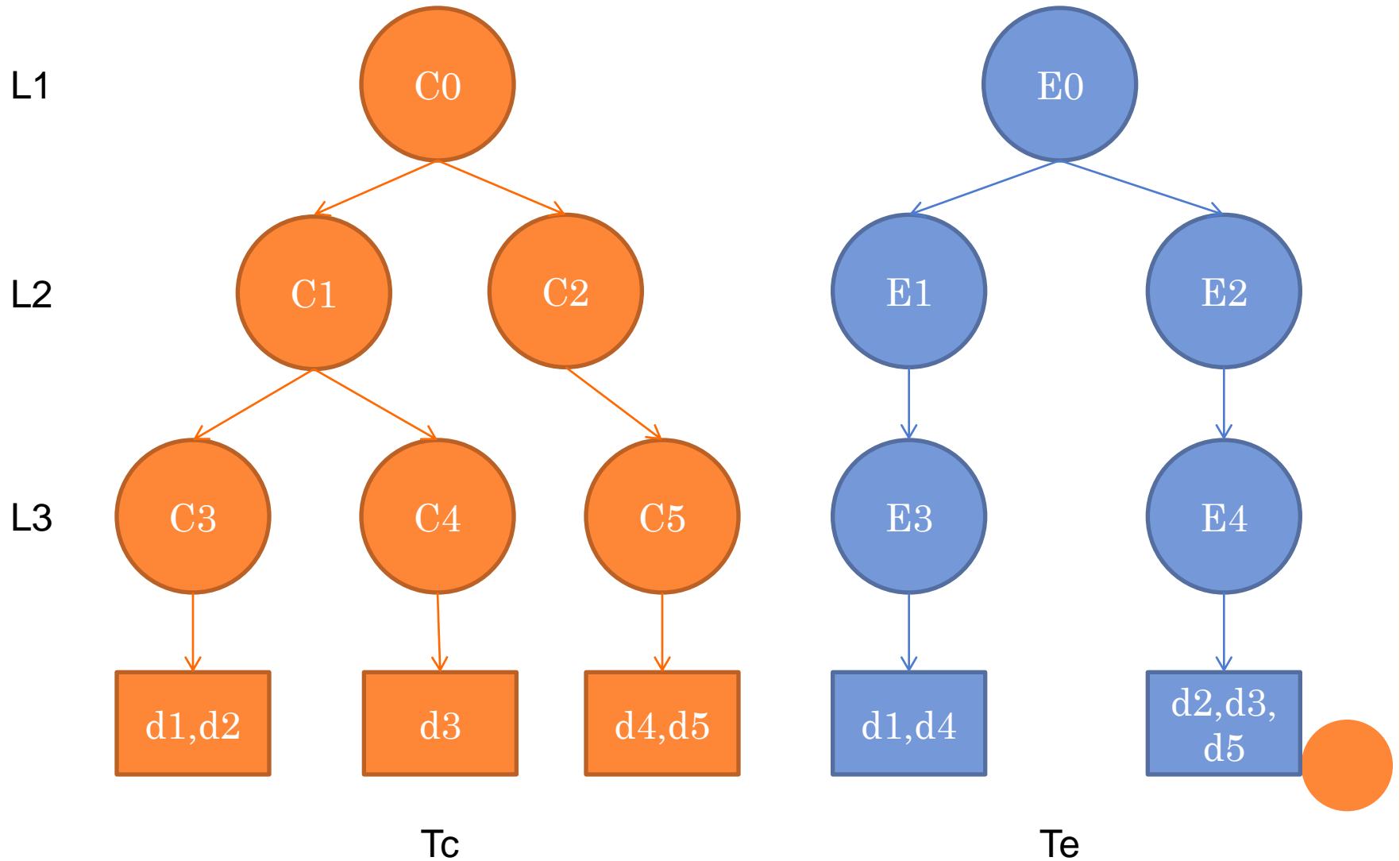


# THE APPROACH OF COMPLETE OPTIMIZED DATA HIERARCHY(CODH)

- Project the CQA data on an external hierarchy, and build a new data hierarchy  $T_e$  by utilizing this new hierarchy.
- Generate a new taxonomy  $T_n$  from  $T_e$  and  $T_c$  by performing split and merge operations top-down on each level
- Note: we get the external taxonomy from Open Directory Project(ODP)

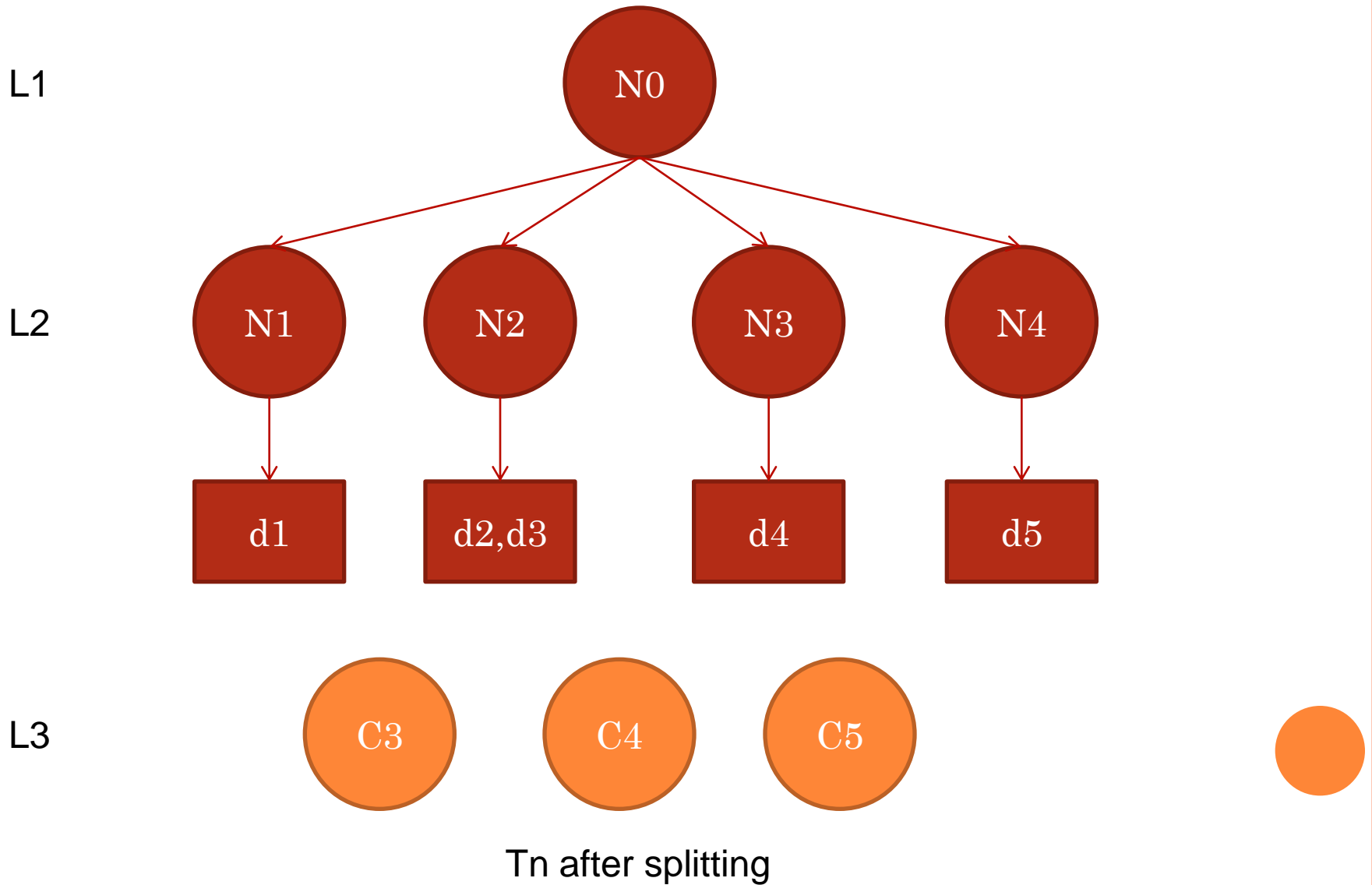


# ORIGINAL TAXONOMIES

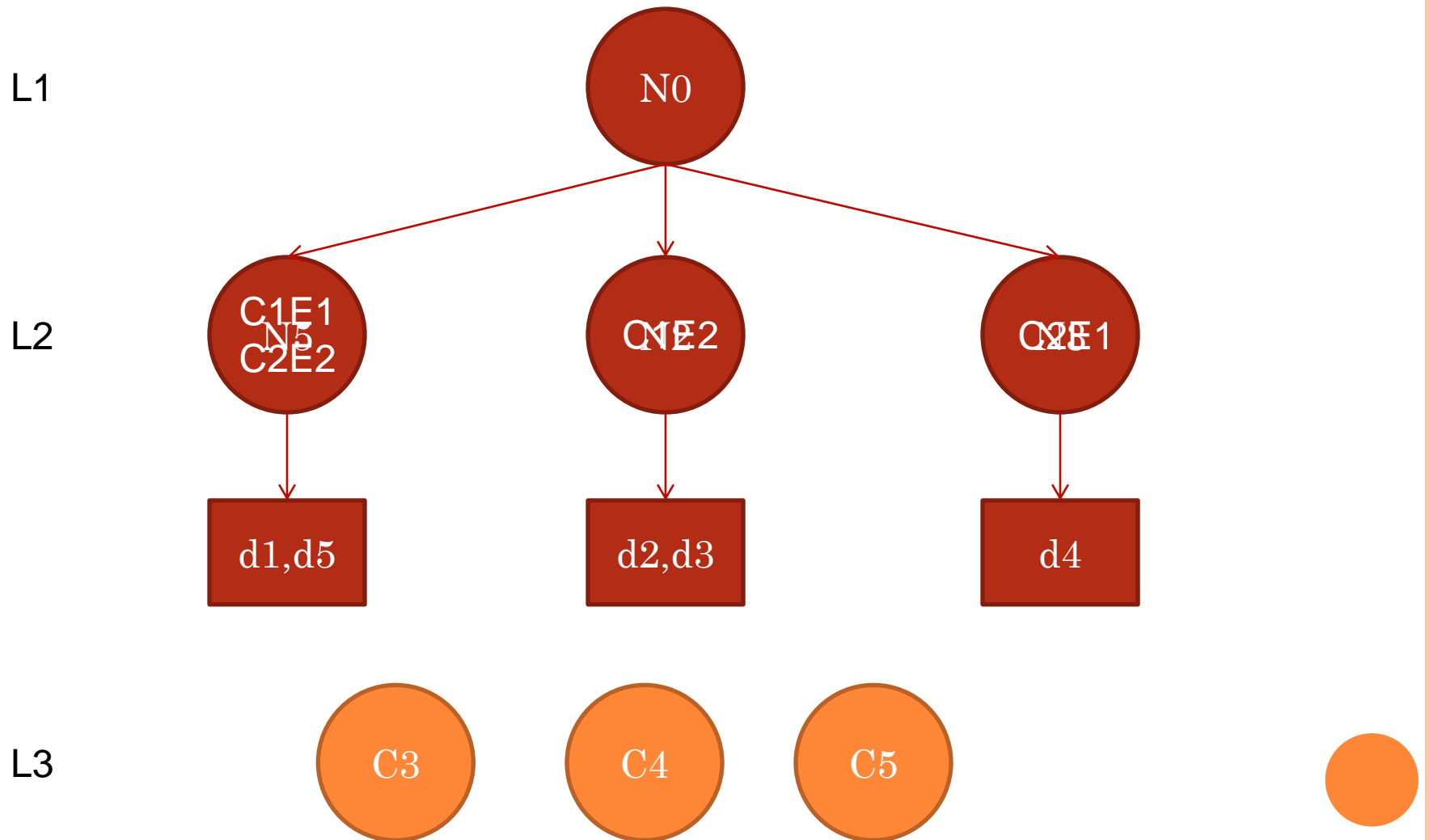




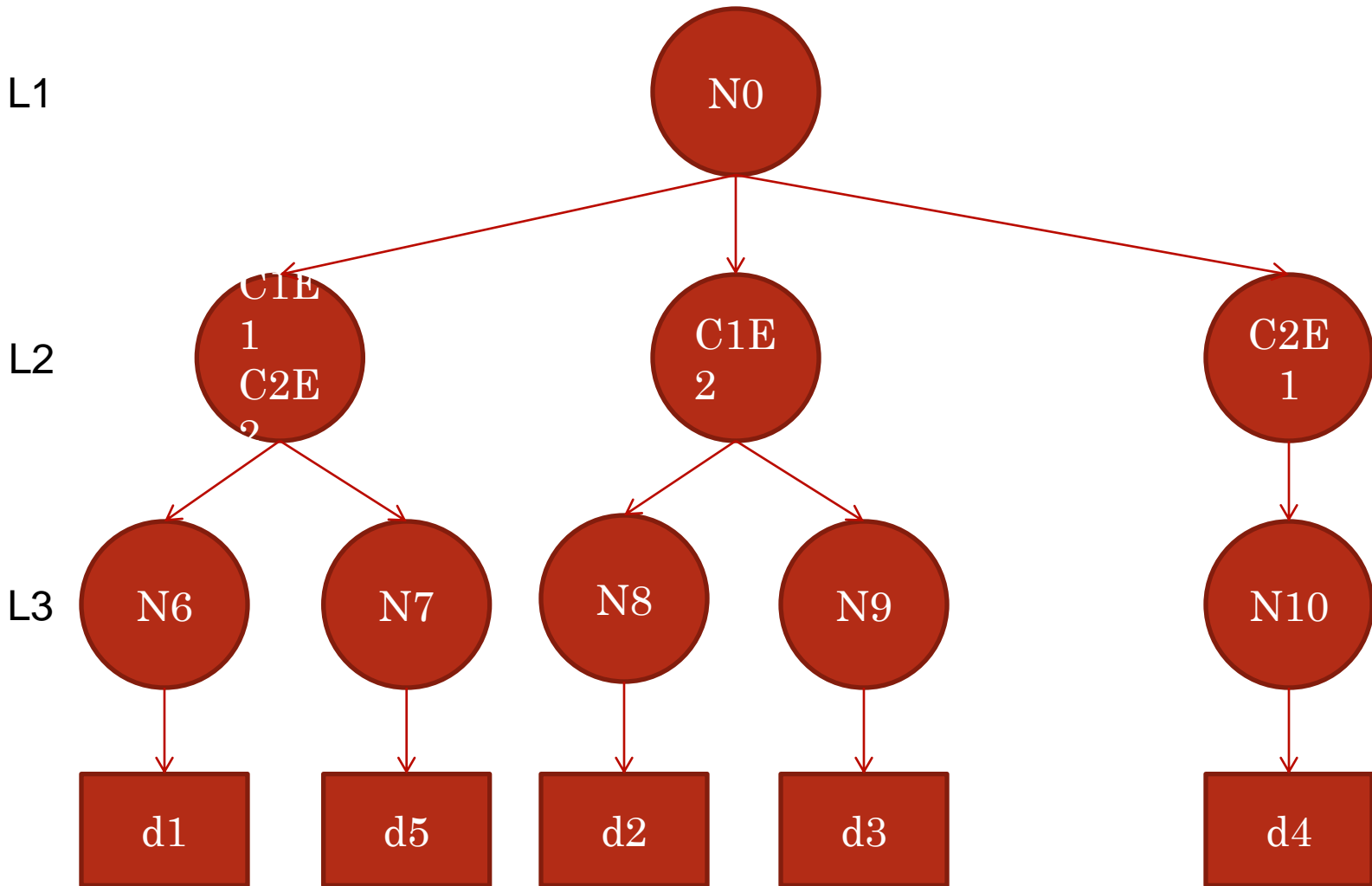
# SPLIT



# MERGE



# UPDATE



# EXPERIMENTS AND RESULTS

## ○ **Classification Performance**

### **Evaluation:**

- 405,814 questions from topics *Computer & Internet* and *Sports* of YA.
- 203,448 descriptions from topics *Computer* and *Sports* of ODP.
- Corresponding hierarchies of YA and ODP.

## ○ **Coherence and Overlap Evaluation:**

- 3,810,739 questions from all topics of YA.
- 1,770,552 descriptions from all of ODP.
- Hierarchies of YA and ODP.

## ○ **Granularity.**



# BASELINE METHODS

- B1: Bottom Up Clustering(BUC)
- B2: Top Down Cluster using Subcategories(TDCS)
- B3: Hierarchical Acclimatization algorithm(HA)
- B4: Question-based Clustering(QC)
- Our approach: CODH

Table 1. Comparison between Algorithms

	BUC	TDCS	HA	QC	CODH
Use $T_c$		partial	√		√
Use $T_e$					√
New leaf				√	√



# CLASSIFICATION AND COHERENCE RESULTS

Table 2. Classification Result

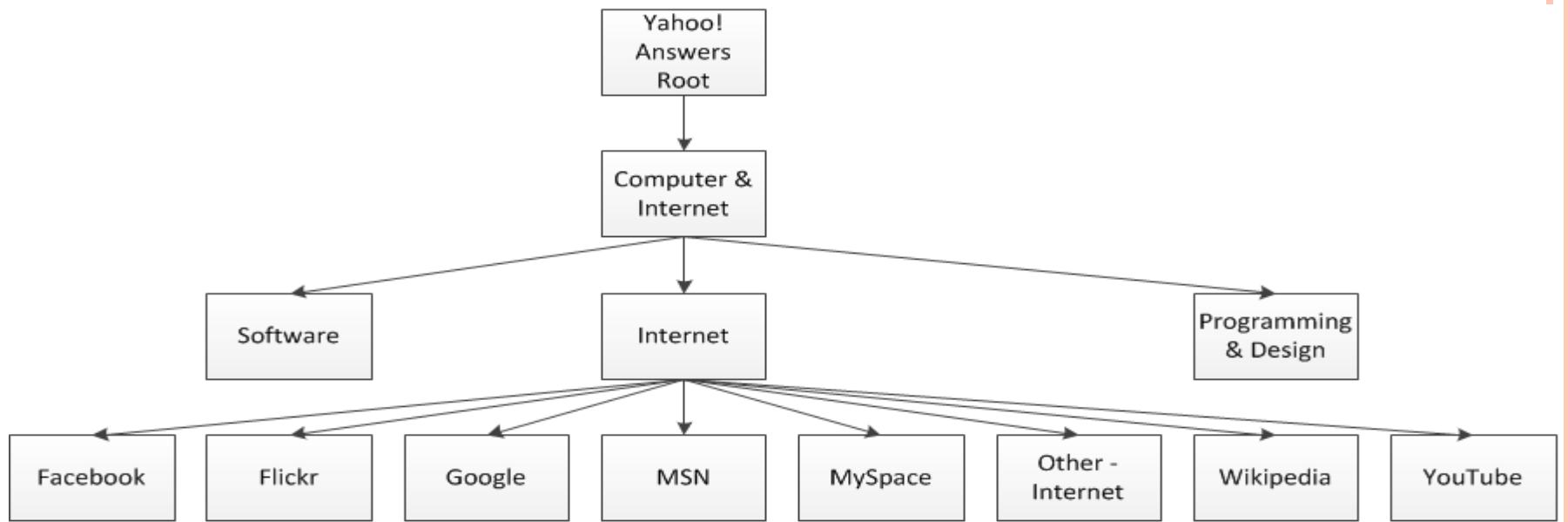
	Original	BUC	TDSC	HA	QC	CODH
Macro F1	0.5188	0.5219	0.4792	0.5340	0.5393	<b>0.7965</b>
Micro F1	0.6391	0.6183	0.5952	0.6336	0.7231	<b>0.8225</b>

Table 3. Coherence Result (The higher the better)

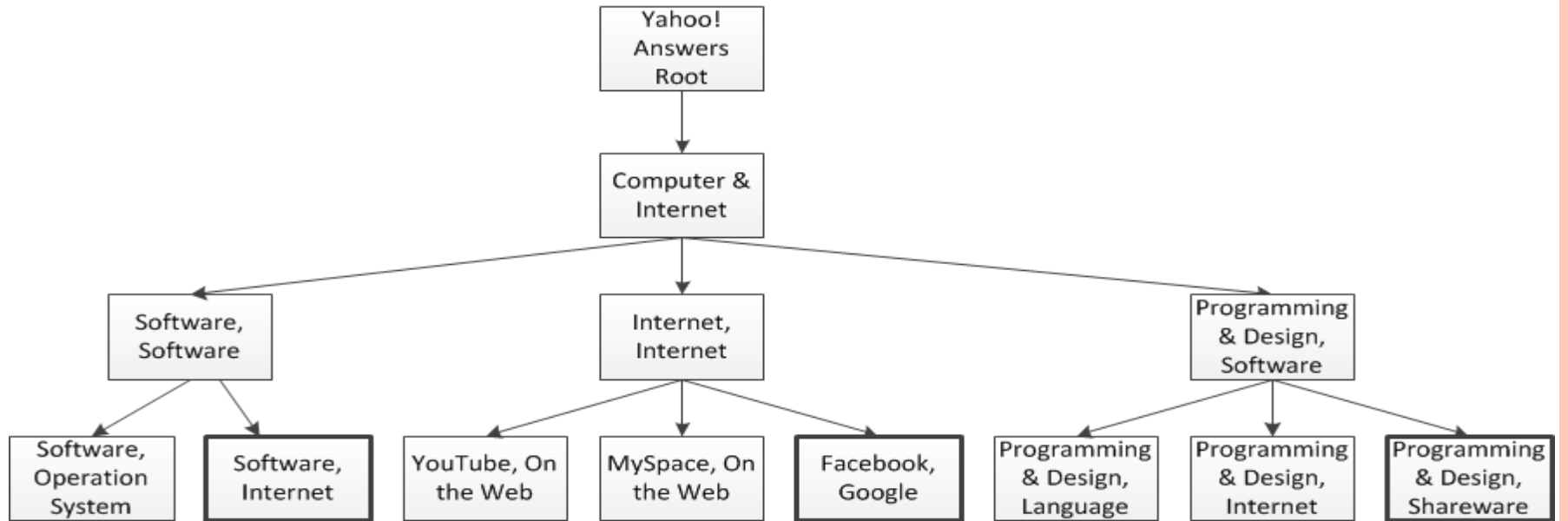
	Original	BUC	TDSC	HA	CODH
Macro F1	0.1242	0.1160	0.1042	0.1312	<b>0.2365</b>
Micro F1	0.3879	0.3673	0.3814	0.3790	<b>0.7592</b>

**CODH outperforms baseline algorithms.**





(T<sub>c</sub>)



(T<sub>n</sub>)

**Thank you!**

