

Collective Intelligence and My Paper

Sohn JongSoo

mis026@korea.ac.kr

Intelligent Information System Lab.

Dept. Of Computer and Information Science

Korea University

Index

- **Concept of collective intelligence**
- **History**
- **Basic Principle**
- **Methodology**
- **Applications**
- **My paper**
- **Discussion about collective intelligence**
- **Conclusion**

Concept of collective intelligence

■ Collective intelligence

- Shared or group intelligence
 - Emerges from the collaboration and competition of many individuals
- A form of networking enabled by the rise of communications technology, namely the Internet.
 - Web 2.0 has enabled interactivity
 - Generating user's own content.

■ Definition

- " Groups of individuals doing things collectively that seem intelligent "
 - *Defined by 'Center for collective intelligent, MIT'*

Concept of collective intelligence

■ Crowd's "collective intelligence"

- Producing better outcomes than a small group of experts (Users add value)
 - Network effects from **user contributions or experiences**
- Requirements for Collective Intelligence
 - Diversity of opinion
 - Independence of members from one another
 - Decentralization
 - A good method for aggregating opinions

■ The study of collective intelligence

- Subfield of sociology, business, computer science, mass communications and mass behavior

Concept of collective intelligence

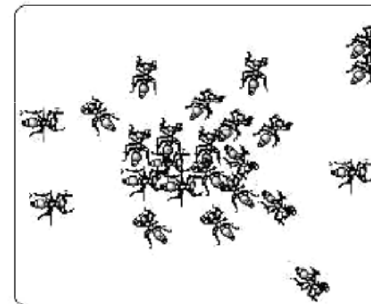
■ Four principles need to exist

- Openness
 - Allowing others to share ideas and bid for franchising
- Peering
 - Horizontal organization with the capacity to create information technology and physical products
- Sharing
 - To expand market and bring products out more quickly
 - Collecting experiences
- Action globally
 - Using standard of web services or technologies

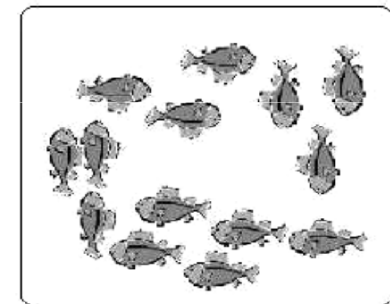
Concept of collective intelligence

■ Collective dynamical behaviors (Couzin, 2002)

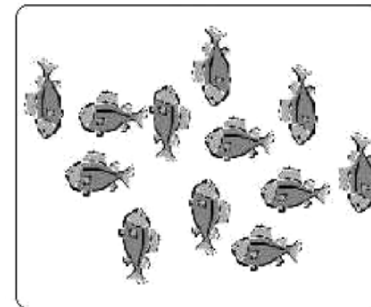
- Swarm
 - Torus
 - Dynamic parallel group
 - Highly parallel group
- Swarm can be viewed as a group of agents cooperating to achieve some purposeful behavior and achieve some goal



(a) Swarm



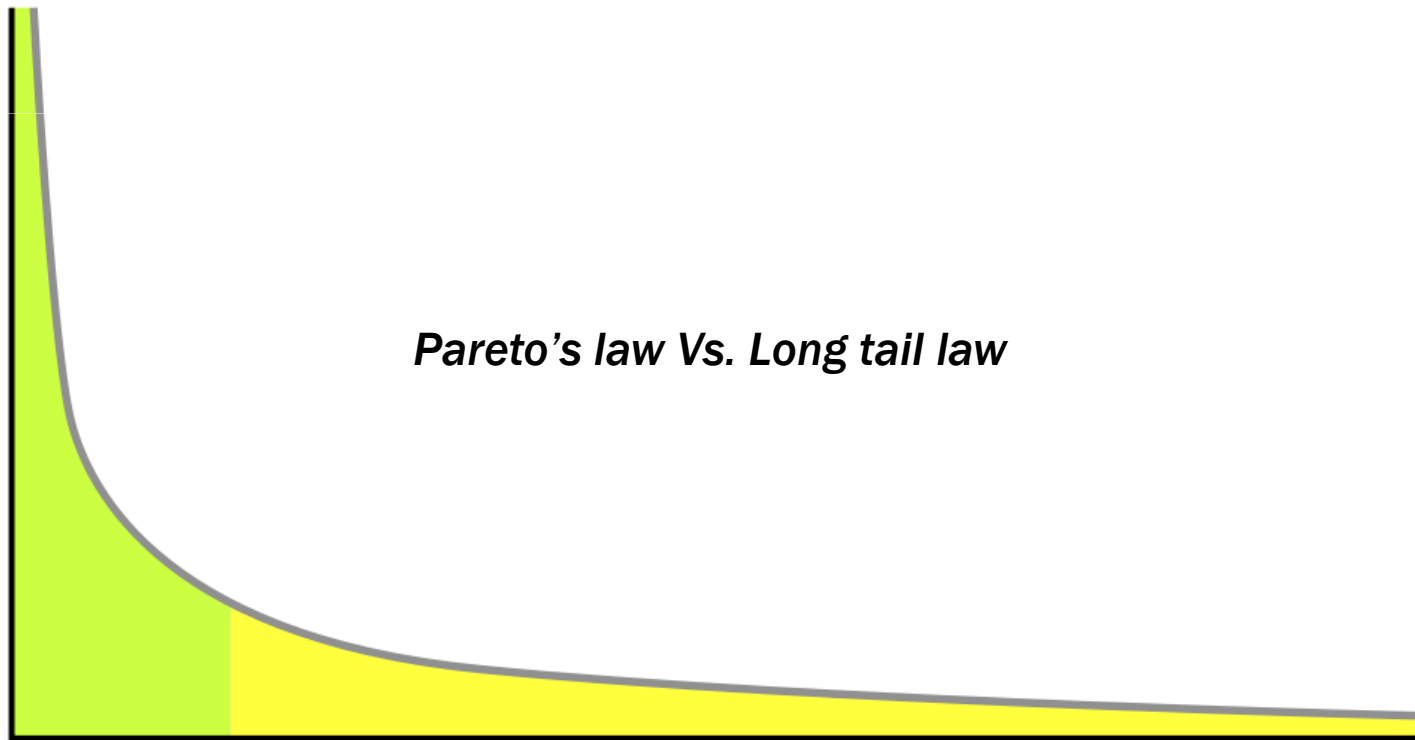
(b) Torus



Concept of collective intelligence

■ Why Do We Care About Collective Intelligence on the Web?

■ Signal vs. Noise in the Long Tail



History

- **William Morton Wheeler, 1911**
 - Collaborative process at work in ants
 - Larger creature that the colony seemed to form a "superorganism"
- **Howard Bloom, 1986**
 - Concepts of apoptosis, parallel processing, group selection and the superorganism
 - Producing a theory of how a collective intelligence works
- **Marco Dorigo, 1992**
 - Ant colony optimization
- **James Kennedy et al, 1995**
 - Particle swarm optimization

History

■ 1990's

- World Wide Web
 - Hypelink
- Environment for collective intelligence

■ Collective behavior (Couzin et al., 2002)

- Homogeneity
 - Every bird in flock has the same behavioral model
- Locality
 - Nearest flock-mates only influence the motion of each bird
- Collision Avoidance, Velocity Matching, Flock Centering

■ Advent of Web2.0

- Attract web engineer's attention
- Using user's experiences and behavior

Basic principle

■ Fact : X_k

- $X_k = a + n_k$
 - a : correct fact (knowledge)
 - n_k : noise

$$x_{CI} = \sum_{k=1}^K x_k = Ka + \sum_{k=1}^K n_k$$

■ Using many facts

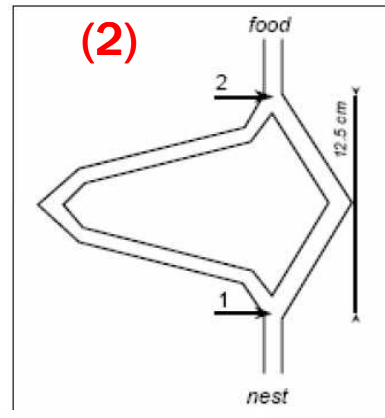
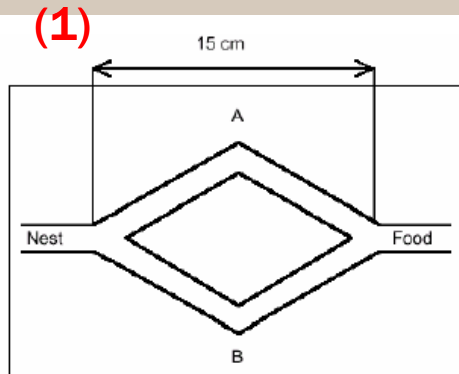
$$x_{CI}[i] = \sum_{k=1}^K \begin{cases} x_k[i_k] & i_k = i \\ 0 & i_k \neq i \end{cases} = K_i a[i] + \sqrt{K_i} n$$

- $X_k[i_k] = a[i_k] + n_k$
 - i -th knowledge of user K

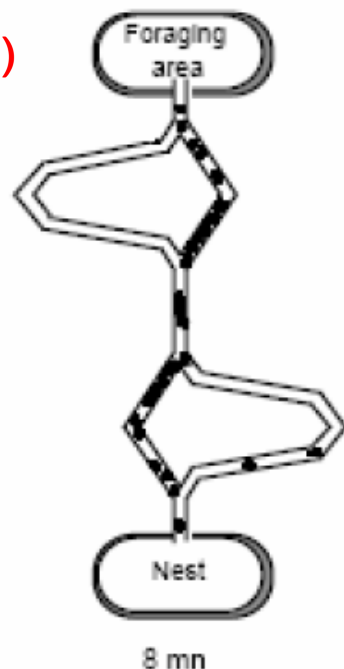
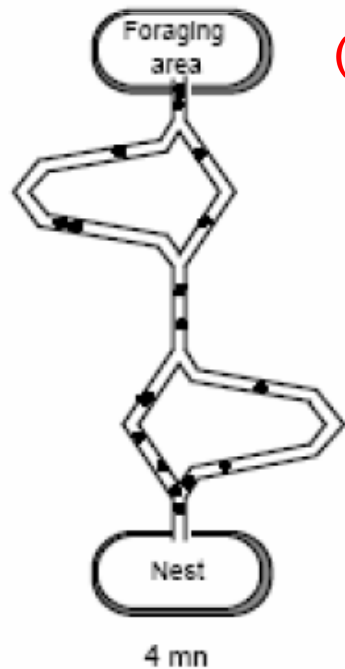
Methodology – ACO

- **Ant colony optimization (ACO)**
 - Inspired by foraging behavior of ants
 - Ants find shortest path to food source from nest
 - Ants deposit pheromone along traveled path which is used by other ants to follow the trail
 - This kind of indirect communication via the local environment is called stigmergy
 - Has adaptability, robustness and redundancy

Methodology - ACO

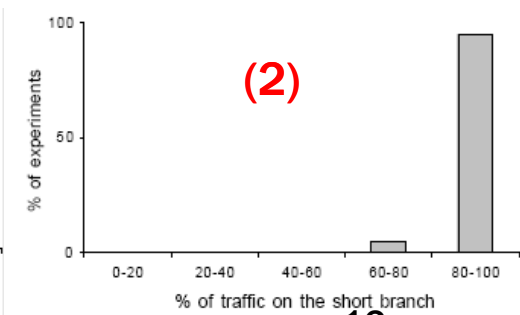
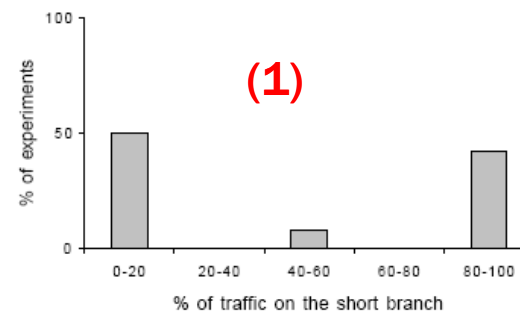


(1) Ants finished all using the same path (each one of the 2 paths, 50% of times)



(2) Ants use the short path

(3) Ants get to find the shortest path



Methodology – ACO

■ Modeling ants colony

- It is known that the ability of ants in finding the shortest route between the nest and a food source can be used to solve **graph problems**.

Graph	Natural Ants	Model
	Individual Ants	Agents
Nodes	Places where ants can stop	Cities
Edges		Routes

- Actions that an agent performs:
 - In a city, it chooses a route based on the intensity of the pheromone over the available paths
 - When it finds the food source, it starts the return travel on its own pheromone trail
- All actions require only **local information** and **short memory**

Methodology – ACO

- Each artificial ant is a probabilistic mechanism that constructs a solution to the problem, using:
 - Artificial pheromone deposition
 - Heuristic information: pheromone trails, already visited cities memory ...

Algorithm 1 Ant colony optimization metaheuristic

Set parameters, initialize pheromone trails

while termination conditions not met **do**

 ConstructAntSolutions

 ApplyLocalSearch {optional}

 UpdatePheromones

end while

Methodology – ACO

■ Update pheromones

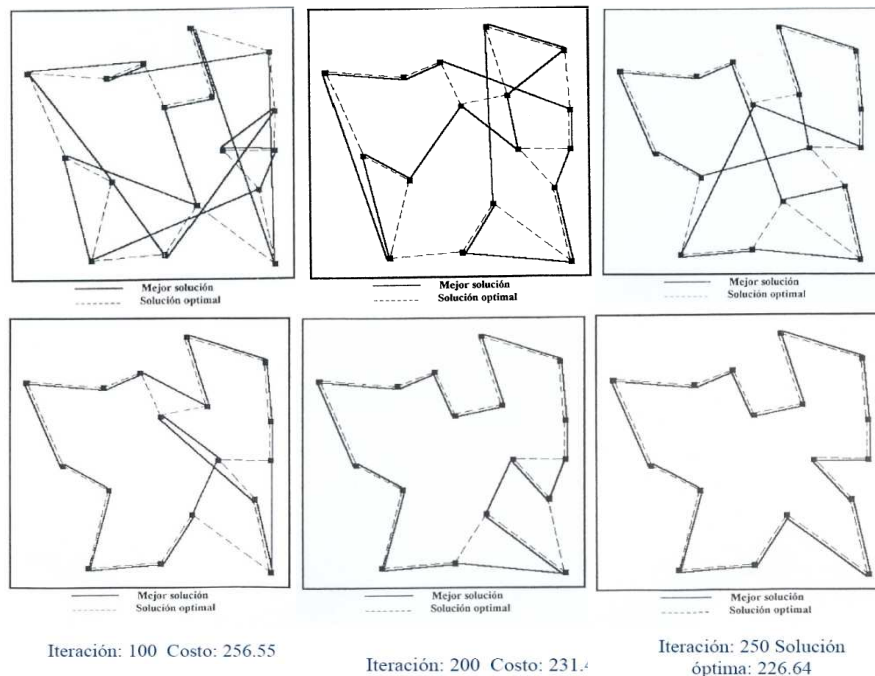
- Increasing the pheromone values
 - Associated with good or promising solutions
- Decreasing the pheromone values
 - Associated with bad ones

$$\tau_{ij} \leftarrow (1 - \rho) \cdot \tau_{ij} + \rho \cdot \sum_{s \in S_{upd} | c_{ij} \in s} F(s)$$

- ▶ S_{upd} is the set of solutions that are used for the update
- ▶ ρ (0; 1] is a parameter called evaporation rate
- ▶ F is a function commonly called the *fitness function*

Methodology – ACO

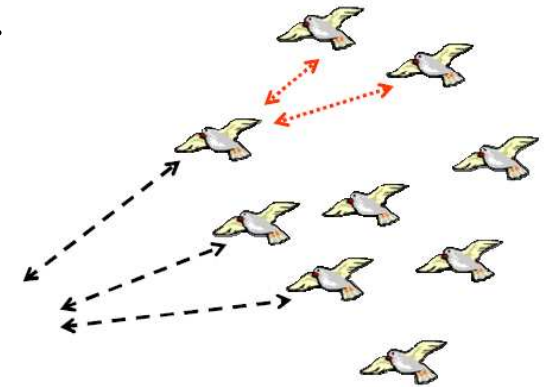
- The ACO can be used to solve graph problems such as the Traveling Salesman Problem (TSP).
 - Of High computational complexity
 - For which the exact algorithms are inefficient
 - For which we don't need *the best* solution but a *good* one.



Methodology – PSO

■ Particle Swarm Optimization (PSO)

- Modeling human or insects social behavior.
- Individuals interact with one another
 - learning from
 - ▶ their own experience
 - ▶ gradually move towards the goal



- Bird flocking is one of the best example of PSO in nature.
- One motive of the development of PSO was to model human social behavior.

Methodology – PSO

■ Algorithm

- Each particle (or agent) evaluates the function to maximize at each point it visits in spaces.
- Each agent remembers the best value of the function found so far by it (**pbest**) and its co-ordinates.
- Secondly, each agent know the globally best position that one member of the flock had found, and its value (**gbest**).

Methodology – PSO

■ Algorithm – phase 1

- Using the co-ordinates of pbest and gbest, each agent calculates its new velocity as:

$$v_i = v_i + c_1 \times \text{rand}() \times (\text{pbestx}_i - \text{presentx}_i) \\ + c_2 \times \text{rand}() \times (\text{gbestx} - \text{presentx}_i)$$

where $0 < \text{rand}() < 1$

$$\text{presentx}_i = \text{presentx}_i + (v_i \times \Delta t)$$

Methodology – PSO

■ Algorithm – phase 2

- In n-dimensional space :

$$\vec{v}_i = \vec{v}_i + \text{rand}() \times \vec{c}_1 \otimes (\overrightarrow{\text{pbest}}_i - \overrightarrow{\text{present}}_i) + \text{rand}() \times \vec{c}_2 \otimes (\overrightarrow{\text{gbest}} - \overrightarrow{\text{present}}_i)$$

cognitive component

social component

Note that the symbol \otimes denotes a point-wise vector multiplication.

Methodology – PSO

Randomly generate an initial population

repeat

for i = 1 to population_size **do**

if $f(\overrightarrow{\text{present}}_i) < f(\overrightarrow{\text{pbest}})$

then $\overrightarrow{\text{pbest}} = \overrightarrow{\text{present}}_i$;

$\overrightarrow{\text{gbest}} = \text{best}(\overrightarrow{\text{pbest}})$;

for d =1 to dimensions **do**

 velocity_update();

 position_update();

end

end

until termination criterion is met.

Applications

■ Amazon recommendation



Customers who bought this item also bought

- [Foundations of Ajax \(Foundation\)](#) by Ryan Asleson
 - [Professional Ajax \(Programmer to Programmer\)](#) by Nicholas C. Zakas
 - [Hibernate in Action \(In Action series\)](#) by Christian Bauer
 - [Agile Web Development with Rails: A Pragmatic Guide \(Pragmatic Programr](#)
 - [Programming Ruby: The Pragmatic Programmers' Guide, Second Edition](#) by
- ▶ [Explore similar items](#) : Books (50)

Similar Books to Explore

you looked at: you might also consider:

[Understanding Search Engines...](#) Paperback by Michael Berry

[Information Retrieval](#) Hardcover by William R. Hersh

[Web Technologies Research and...](#) Paperback

These phrases occur frequently in each of these books: "document vectors" ([learn more](#))

What do customers ultimately buy after viewing this item?

- 75%** buy the item featured on this page: [Ajax in Action](#) by [Dave Crane](#) ★★★★★ \$29.67
- 9%** buy [Professional Ajax \(Programmer to Programmer\)](#) by [Nicholas C. Zakas](#) ★★★★★ \$26.39
- 6%** buy [Head Rush Ajax](#) by [Brett McLaughlin](#) ★★★★★ \$26.39
- 6%** buy [JavaScript: The Definitive Guide](#) by [David Flanagan](#) ★★★★★ \$32.99
- 4%** buy [CSS Mastery: Advanced Web Standards solutions](#) by [Andy Clark](#) ★★★★★ \$23.99

[Compare these items](#) [Explore Similar Items](#)

So many Recommendations!

Applications

■ <http://del.icio.us/tag>

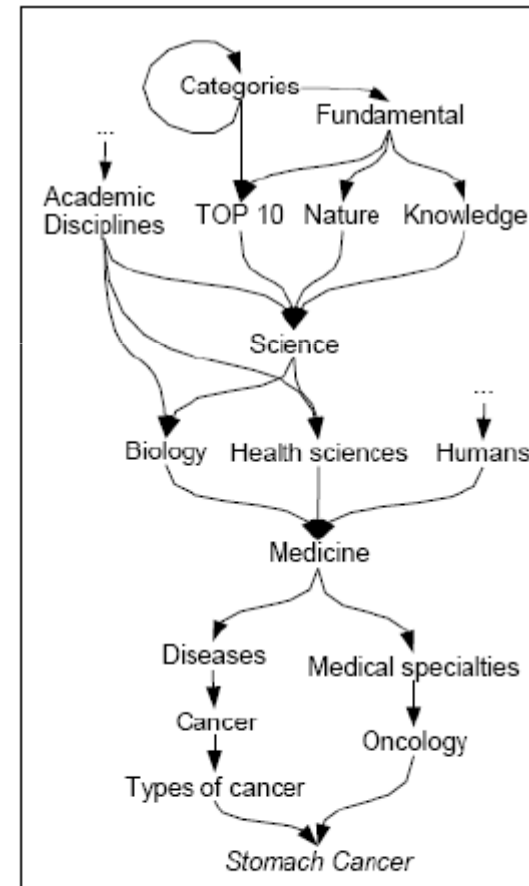
- Practice and method of collaboratively creating and managing tags to annotate and categorize content
- collaborative tagging, social classification, social indexing, and social tagging

.net advertising ajax apple architecture **art** article au audio **blog** blogs books business
collaboration comics community computer cooking cool crafts **CSS** culture **design** development diy
download drupal **education** email environment fashion fit film finance **flash** food free
freeware fun funny **games** google government graphics green gtd hardware health history home
howto html humor illustration images imported india **inspiration** interactive interesting
internet iphone java javascript jobs language learning library lifehacks **linux** literature mac maps
marketing math media microsoft mobile money movies mp3 **music** news online opensource
osx photo **photography** photos photoshop php politics portfolio productivity
programming psychology python rails reading recipe **recipes** reference research resource
resources ruby science search security seo sga shop **shopping** slash social **software**
teaching tech **technology** tips tools toread travel **tutorial** tutorials tv typography
ubuntu **video** videos web web2.0 webdesign wedding wiki windows wordpress work
writing youtube

Applications

■ Thesaurus tagging

- Collaborative thesaurus tagging the Wikipedia
 - *Jakob Voss, 2006*



from the English Wikipedia

Applications

■ Google page rank

- The Philosophy of PageRank
- Uniquely democratic nature of the web
 - By using its vast link structure as an indicator of an individual page's value
 - Interpreting a link from page A to page B as a vote, by page A, for page B



PageRank Algorithm

$$PR(A) = 1 - d + d \left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \dots \right)$$

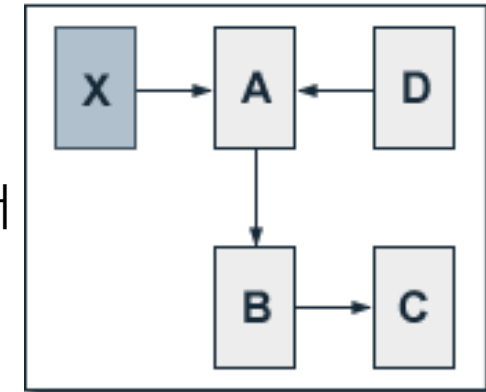
- PR(A) : PageRank of page A
- L(A) : The number of links going out of page A
- d : damping factor

Applications

■ Google page rank

■ Damping factor

- 랜덤 서퍼가 어떤 페이지를 읽다가 지루해져서 또 다른 랜덤 페이지를 찾게될 확률
- 통상적으로 0.85
 - ▶ 선택적으로 조절



■ Example

- $d = 0.5, PR(D) = PR(X) = 5$
- $PR(A) = 0.5 + 0.5 ((PR(X) / 1) + (PR(D) / 1)) = 5.5 + 0.5 PR(D)$
- $PR(A) = 5.50, PR(B) = 3.25, PR(C) = 2.10$

- $d = 0.85, PR(D) = PR(X) = 5$
- $PR(A) = 8.65, PR(B) = 7.50, PR(C) = 6.52$

My paper

■ 집단지성을 이용한 온톨로지 생성 시스템의 설계

- 온톨로지
 - 다양한 온톨로지 생성 방법 및 도구 등장
 - ▶ 작성이 어렵고 난해함
 - ▶ 소수의 전문가 집단
- 웹2.0
 - 콘텐츠의 생산주체
 - ▶ 서비스 제공자에서 수요자로 변화
- 집단지성을 이용한 온톨로지 생성
 - 사용자 생성 온톨로지를 사용
 - 집단지성의 개념을 도입하여 협업적 온톨로지 생성
 - ▶ 원시 온톨로지 -> 매개 온톨로지 -> 대중 온톨로지
 - 기존 온톨로지 생성의 단점을 효과적으로 해결

My paper – 배경

■ 시맨틱 웹 및 온톨로지

■ 시맨틱 웹

- 팀 버너스 리 (2001')
- 정보자원의 의미를 컴퓨터가 스스로 인지
 - ▶ 자동적으로 데이터를 처리
- 온톨로지
 - ▶ 분산 컴퓨팅 환경에서 지식의 명시적인 명세

■ 온톨로지 언어 및 생성

- RDF, DAML+OIL, OWL 등의 언어
- 특정한 분야에 한정된 사용
- 온톨로지 생성의 어려움
 - ▶ 생성 방법이 복잡함
 - ▶ 온톨로지 생성 자동화가 어려움

My paper – 배경

■ 웹2.0

- 닷컴 기업들의 붕괴 이후 살아남은 기업들의 특징 분류
 - O'Reilly Media – 데일 도허티
- 콘텐츠 생산 주체의 변화
 - 서비스 운영자에서 사용자
 - 집단지성(Collective intelligence)의 활용
 - ▶ 다수의 사용자가 신뢰성 있는 지식을 축적
 - ▶ Wisdom of crowds, Swarm intelligence
- 데이터의 정형화
 - XML 기반 데이터 포맷
- 서비스 지향적
 - REST, SOAP 등의 서비스 프로토콜
- 매시업 (Mash-up) 어플리케이션
 - 서비스 제공자가 제공하는 API를 이용한 어플리케이션
 - 플러그인(Plug-in)과 비슷한 개념

My paper – 기존 온톨로지 생성 방법 및 도구

■ OTKM

- **OnTo Knowledge Methodology**
 - for development and evolution of Ontology-based Tools for Knowledge Management
- 어플리케이션 지향 온톨로지 개발에 초점
- 온톨로지 생성에 방대한 시간 및 비용 소요

■ ONIONS (Ontologic Integration On Naive Sources)

- 도메인 온톨로지를 분석 통합을 목적

■ Protege

- GUI기반 온톨로지 제작도구
- 온톨로지를 사용자가 일일이 제작

My paper – 집단지성을 이용한 온톨로지 생성

■ 시스템 설계의 요점

- 사용자들이 어렵지 않게 개념을 도식화하고 활용할 수 있어야 함
- 시스템 설계의 요점 두가지
 - 접근의 용이성
 - ▶ Open API의 제공
 - 블로그, 위키, 팟캐스트 등의 2.0 어플리케이션에서 매쉬업 어플리케이션을 제작할 수 있도록 Open API가 제공되어야함
 - 사용자들이 온톨로지 생성 시 접근의 용이성 제공
 - 사용의 편의성
 - ▶ 일반 사용자들이 사용하기에 상대적으로 어려운 개념의 온톨로지
 - ▶ 클래스 관계 표현 및 속성 표현을 최소화
 - 상 하위 관계, 동치, 반의 관계로 한정
 - 제한(Restriction)의 제외

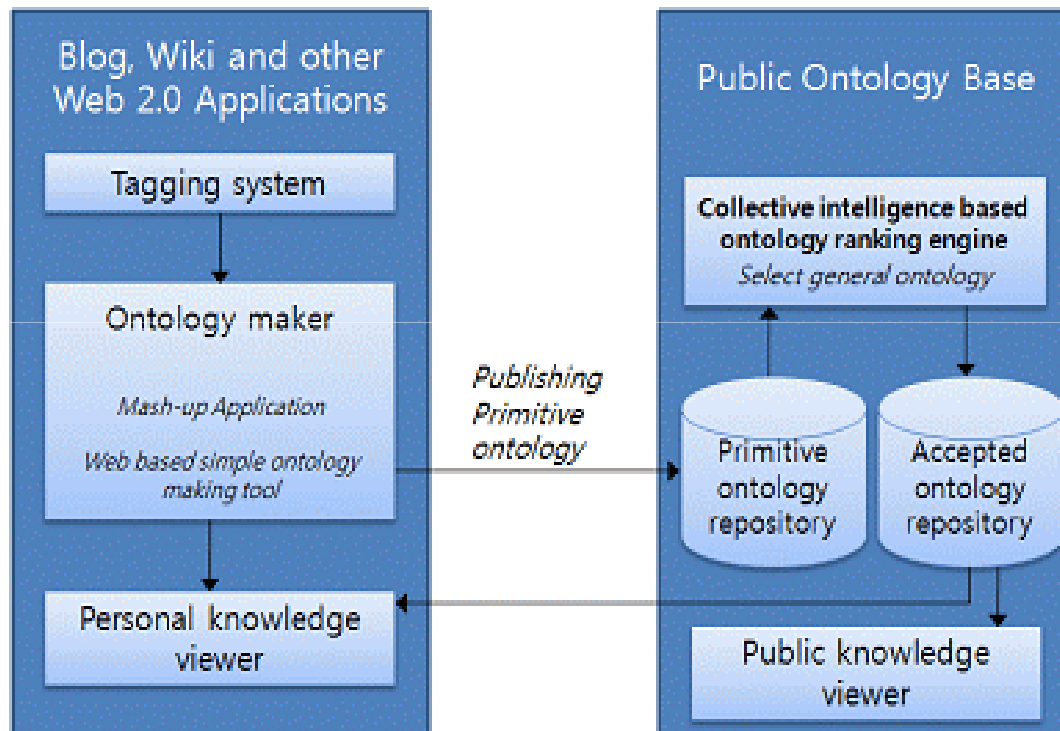
My paper – 집단지성을 이용한 온톨로지 생성

■ 용어의 정의

- 원시 온톨로지
 - 사용자들이 웹2.0 어플리케이션에서 제작한 온톨로지
 - 사용자가 만들어서 서버로 전송
- 매개 온톨로지
 - 원시 온톨로지를 통합한 중간형태의 온톨로지
 - 사용자들에게 전송받은 원시온톨로지를 서버에서 통합
- 대중 온톨로지
 - 집단지성을 이용하여 매개온톨로지를 가공한 형태의 온톨로지
 - 서버에서 생성하여 사용자에게 전송

My paper – 집단지성을 이용한 온톨로지 생성

■ 시스템 아키텍처



My paper – 집단지성을 이용한 온톨로지 생성

■ 온톨로지의 생성 방법 및 절차

■ 집단지성

- 개미가 집단의 의사활동을 이용하여 최적의 경로를 탐색하는 현상을 분석
- 페르몬을 이용하여 다른 개미들과 의사소통
 - ▶ 페르몬 밀도에 따라 긍정과 부정을 구분

■ 집단지성의 적용 모델

[14]의 모델을 활용

J.M. Cueva Lovelle et al. Swarm Intelligent Surfing in the Web" ICWE 2003, LNCS 2722, pp. 431-440, 2003

대상	Lovelle et al.	본 논문
곤충 사회	웹 시스템	웹 시스템
개미	웹 클라이언트/서버	웹 2.0 유저
음식	웹에 산재한 정보	잘 정의된 온톨로지
음식 구하기	웹 서핑	온톨로지 취득
페르몬 분비	웹 서버의 로그 데이터	원시 온톨로지 생성
상호작용	클라이언트/서버간의 요청 및 응답	온톨로지 통합
페르몬 밀도	웹 페이지의 대중성 (Popularity) 및 중요성	온톨로지의 대중성

My paper – 집단지성을 이용한 온톨로지 생성

■ 온톨로지의 생성 방법 및 절차

- 대중 온톨로지의 생성 절차
 - Step1) 원시 온톨로지의 생성
 - ▶ 각 사용자들이 원시 온톨로지를 생성
 - ▶ 온톨로지 저장소로 온톨로지를 전송
 - Step2) 원시 온톨로지의 취합
 - ▶ 전송받은 원시 온톨로지를 취합
 - ▶ 동일한 개념에 대한 온톨로지 군집화
 - Step3) 매개 온톨로지의 생성
 - ▶ 동일 개념에 대한 빈도 계산 및 통합
 - Step4) 대중 온톨로지의 생성
 - ▶ 매개 온톨로지의 RDF 트리플 중에서 높은 페르몬 값을 갖는 노드를 선택하여 대중 온톨로지를 생성

My paper – 집단지성을 이용한 온톨로지 생성

■ 원시 온톨로지의 취합

- 하나의 원시온톨로지 PO, 취합한 원시 온톨로지 집합 S
 - $S = \{PO_1, PO_2, \dots, PO_n\}$
- PO는 RDF 트리플 D의 집합
 - $D = \{(s, r, o) \mid (s, r, o) \in PO\}$

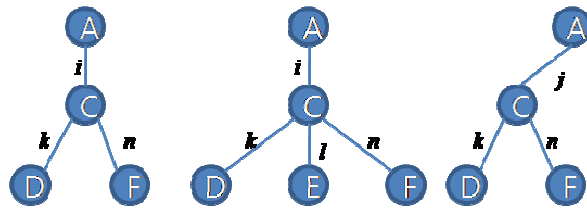
■ 빈도 계산 (페르몬 계산)

- 페르몬 p, 페르몬 밀도 V, 마지막 접근 시간 T
 - $(p : D \rightarrow V \times T)$
 - ▶ 페르몬 밀도 - $P_v(d)$
 - 범위 : 0 ~ 1
 - ▶ 접근시간 - $P_t(d)$
 - 범위 : 1 ~ 2
- S에 속한 어떤 subject S_k 를 갖는 D들의 집합

$$p_v(d) := \frac{\text{number of } D}{\text{number of entities in } DS_k}$$

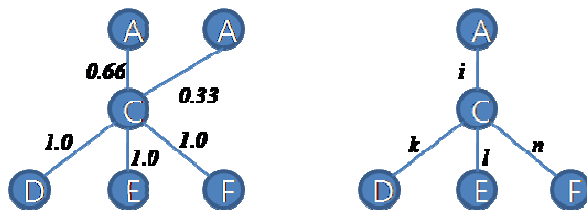
My paper – 예제 및 평가

■ 원시 온톨로지 PO₁, PO₂, PO₃



- {A, i, C} – D1, D2 {A, j, C} – D3
 - $Pv(d) = 0.66$ $Pv(d) = .033$
 - ※ $Pt(d)$ 는 1로 고정

■ 매개 온톨로지와 대중 온톨로지



- {A, j, C} 제거

My paper – 예제 및 평가

■ Protege와의 비교

항목	Protege	집단지성
생성 자동화	반자동	반자동
생성주체	전문가	비전문가
참여자	소수	불특정 다수
GUI	지원	지원
문법	OWL-DL	RDF/OWL-lite
제작난이도	중	하
추론	가능	불가능

My paper - 결론

■ 집단 지성의 활용

- 집단 지성을 이용한 온톨로지 생성 시스템
 - 집단지성을 이용한 온톨로지 생성 시스템 아키텍처를 제시
 - 온톨로지 생성 방법을 제안
- 웹2.0의 중요한 요소 중 한가지인 집단지성을 활용
 - 온톨로지 생성주체의 변화
 - ▶ 소수의 전문가 집단 -> 다수의 일반 사용자
- 대중성있는 대량의 온톨로지를 구축 가능
- 시맨틱 웹과 웹2.0의 접목

■ 향후 연구 과제

- 개념화시킨 온톨로지 생성방법을 실체화
- 사용자들이 사용할 수 있도록 공개형으로 개발 필요
- 계산 모형을 보다 구체화시킬 필요 있음
 - 예외상황 및 세분화된 처리를 강화

Discussion about CI

■ Weak and strong points

- Weak point
 - Hard to understanding computing model
 - ▶ Biological terms + statistical knowledge + linear algebra + various clustering algorithm + etc
 - Hard to establish computing model for problem solving
 - Unexplored field
- Strong point
 - Unexplored field
 - Distributed computing environment
 - ▶ Useful for agent system
 - ▶ Routing
 - ▶ Knowledge processing, datamining
 - Computing cost is very cheap!

Discussion about CI

■ How to adapt collective intelligence to Web?

- Swarm methodology
 - ACO, PSO
- Optimization algorithm aiding
 - Euclidean distance, K-means algo., etc

■ Where we adapt collective intelligence?

- Ontology creating
- Web document classification
- Making thesaurus
- Information retrieval
- Intrusion detection
- Network routing

■ Can we make something new?

- Various parallel algo. , agent technique

Conclusion

- **Rising technique**
 - Web2.0
 - Web service
- **In the distributed computing environment**
 - Low computing cost
 - From the agent point of view
- **Using user's experiences**
 - To make robust web
 - Mass classification

Conclusion

■ My impression

- 생각보다 오랜 역사
 - 1900년대 초반에 첫 개념이 등장
- 이론(Theory)의 단계에 이르지 못하는 것
 - 개념과 방법론은 다양하게 제시
 - 그러나 하나의 이론으로서 일관성있는 모델이 제시되지 않는 것
- 집단지성의 핵심
 - 기존의 알고리즘 및 자료구조에 사용자의 "경험"을 포함
 - ▶ 다양한 알고리즘에 접목 가능
 - ▶ 인공지능망과의 접목 가능성
- 다양한 분야에서 연구 시도 중
 - 웹, 시맨틱 웹, 기업정보시스템, E-learning, 네트워크 등
 - ▶ 2007년을 기점으로 2008년에 다양한 아이디어 제시
- 마음이 급해짐과 동시에 흥미를 느꼈음

Thank you