

ACO 를 이용한 협업적 온톨로지 생성 방법

손중수, 권경락, 정인정
고려대학교 컴퓨터정보학과
e-mail: {mis026, helpnara, chung}@korea.ac.kr

Cooperative Ontology Generation Method Using ACO

Sohn Jongsoo, Kwon Kyunglak, InJeong Chung
Dept of Computer and Information Science, Korea University

요 약

온톨로지는 시맨틱 웹의 핵심 기술로써 시맨틱 웹이 소개된 이후 다양한 온톨로지 생성 방법의 연구가 이루어져 왔다. 그러나 온톨로지는 작성이 어렵고 난해한 면이 있어 소수의 전문가 집단에 의해서만 만들어지고 있는 것이 현실이다. 본 논문에서는 웹 2.0 기반 환경에서 사용자들이 생성한 온톨로지를 수집하여 대중 온톨로지를 완성하는 새로운 온톨로지 생성 방법을 제안한다. 본 논문에서는 집단지성을 이용한 최적화 기법 중 한가지인 ACO 를 온톨로지 생성 분야에 적용시켜 전문가가 아닌 일반 사용자들이 작성한 낮은 수준의 온톨로지를 모아 완성된 형태의 온톨로지를 생성한다. 그리고 본 논문에서 제안한 방법을 통해 만들어진 온톨로지의 신뢰성을 검증하기 위하여 전문가 집단이 만든 온톨로지에 대해 정확도와 재현율을 계산하여 보인다. 본 논문에서 제시하는 방법은 복잡하고 난해한 기존 온톨로지의 제작 방법의 단점을 효과적으로 해결하며 대중적으로 시맨틱 웹이 활용될 수 있는 환경을 구축할 수 있다.

1. 서론

시맨틱 웹은 전통적인 웹의 문제를 극복하기 위하여 팀 버너스 리에 의해 제안되었다[3]. 이에 따라 시맨틱 웹을 구축하기 위하여 국내외에서 다양한 온톨로지 생성 방법론이 제안되고 심도 있는 연구가 이루어지고 있다. 그러나 온톨로지의 제작이 복잡하고 다양한 도메인의 구축된 온톨로지가 부족하여 시맨틱 웹이 빠른 속도로 확산되고 있지는 않다.

한편, 웹 2.0 이 제시되면서 웹 기반 서비스의 환경이 빠르게 변하고 있는데 그 중의 한가지로서 콘텐츠 생산 주체의 변화를 들 수 있다[1]. 이전의 웹 환경에서 인터넷 콘텐츠의 생산 주체가 서비스 운영자였다면 웹 2.0 에서 콘텐츠의 생산 주체는 사용자이다. 이에 따라 사용자 제작 콘텐츠를 분류하고 순위를 매기는 방법들이 다양하게 고안되었는데 이 방법들에 집단지성이 사용 되어지고 있다.

이에 따라 본 논문에서는 시맨틱 웹 구축을 위해 필수불가결한 온톨로지를 웹 2.0 의 주요 기술 중 한가지인 집단지성을 이용하여 사용자가 직접 생성하고 공유하는 방법을 제안한다. 본 논문에서 제안하는 온톨로지 생성 방법은 집단지성을 기반으로 하는 최적화 알고리즘 중 하나인 ACO(Ant Colony Optimization)[18]를 활용한다. 본 논문에서 제안하는 방법은 크게 1) 원시 온톨로지의 생성, 2) 매개 온톨로지의 생성 그리고 마지막으로 3) 대중 온톨로지의 생성, 이렇게 세 온톨로지의 생성 과정을 통해 온톨로지를 생성한다. 본 논문에서 제안하는 온톨로지 생성 방법은 비록 자동적인 생성방법은 아니지만 다수의

사용자가 큰 노력을 들이지 않고 협업적으로 온톨로지 생성이 가능하며 개개인의 온톨로지를 공유하고 공동의 온톨로지를 추출함으로써 대중적으로 사용 가능한 온톨로지 집합을 구축할 수 있다.

2. 관련 연구

시맨틱 웹은 기존의 월드와이드웹을 확장한 형태로써 컴퓨터가 이해할 수 있는 형태로 정보자원의 의미 및 관계를 기술하여 컴퓨터와 사람이 협동적으로 업무를 수행할 수 있는 형태의 웹을 지칭한다.[3] 시맨틱 웹을 구현하기 위하여 정보자원에 대한 의미기술이 필수적인데 이 때 사용되는 것이 온톨로지이다. 그러나 무한대에 가까운 모든 사물에 대하여 온톨로지를 기술하는 것은 어렵고 복잡한 문제이므로 현재 온톨로지는 특정한 분야에 대해 한정되어 사용되고 있다. 그리고 그 구축 방법의 까다로움으로 인해 시맨틱 웹이 확산되는데 어려움으로 작용하고 있다. 이에 따라 온톨로지를 자동적 혹은 효율적으로 생성하기 위하여 다양한 연구가 이루어지고 있다.

온톨로지를 생성하기 위한 저작 도구로는 Protégé[6][7], OntoEdit[10] 등이 있다. 이 저작 도구들은 모두 GUI 기반 도구로써 다양한 분야의 온톨로지를 생성하고 시각화할 수 있는 환경을 제공한다[13]. 그러나 GUI 기반 온톨로지 저작도구들은 온톨로지들을 사용자들이 일일이 작성해야한다는 단점이 있으며 다양한 분야의 온톨로지를 대량으로 생성하기 어렵다.

온톨로지를 생성하기 위한 방법으로는 OTKM[4], ONIONS[5]등이 있다. 이들 방법들은 온톨로지를 생

성하고자 하는 분야의 지식들을 개념화의 단계를 거쳐 개념을 정제하고 지식의 상관성을 분석하여 온톨로지를 생성하기 위한 방법을 제안한다. 이 방법들은 특정한 분야의 온톨로지를 상세히 생성하는데 적절한 방법이지만 최근의 폭발적인 웹 정보자원의 증대와 다양한 분야의 지식을 다루기에는 한계가 존재한다.

한편, [19]에서는 집단지성을 활용하여 많은 사용자들이 협업적으로 온톨로지를 생성하기 위한 시스템을 제안했으나 온톨로지를 생성하는 구체적인 방법에 대한 제시가 없이 개념적인 구조만을 보였다.

3. ACO 를 이용한 협업적 온톨로지 생성 방법

3.1. ACO (Ant Colony Optimization)

ACO 는 개미들이 페르몬을 이용하여 어떤 시작점에서 목표점까지 가장 빠른 길을 협업적으로 찾아가는 과정을 정형화한 집단지성 기반 휴리스틱 알고리즘이다[18]. 시작점과 목표점을 연결하는 최단거리 문제를 해결하기 위해 초기해에 해당하는 하나의 경로를 찾은 후 여러 회를 반복하면서 전 단계에서 남겨진 값에 일정 값을 곱하고 페르몬의 증발률을 반영하여 최종적인 페르몬 값을 계산하는 과정을 반복한다. ACO 에서는 어떤 꼭지점 i 에서 다른 꼭지점 j 사이에 남겨진 페르몬을 τ_{ij} 로 표기하며 ρ 가 증발률, m 이 개미의 숫자라고 할 때 다음 (식 1) 같이 계산된다.

$$\tau_{ij} \leftarrow (1-\rho) \cdot \tau_{ij} + \sum_{k=1}^m \Delta \tau_{ij}^k$$

$$\Delta \tau_{ij}^k = \begin{cases} \frac{\rho}{L_k} & k \text{ 번째 개미가 꼭지점 } (i,j) \text{ 를 여행했을 때} \\ 0 & \text{그 외의 경우} \end{cases}$$

(식 1)

그리고 페르몬의 변화량은 개미 k 가 i, j 에 해당하는 길을 갔을 때 두 꼭지점 사이의 거리에서 상수 ρ 를 나눈 값으로 계산된다. 본 절에서 보인 ACO 는 아래 3.2.절에서 대중 온톨로지를 생성하기 위해 페르몬을 계산하는데 활용된다. 즉, ACO 에서 페르몬의 업데이트는 새로 추가된 페르몬에 $(1 - \text{증발률})$ 을 곱한 값에 누적된 페르몬 값을 더한 값으로 결정되는데 본 논문에서는 이 계산 모형을 기초로 하여 어떤 두 클래스에 대해 특정 관계를 기술한 원시 온톨로지의 기술한 비율에 따라 페르몬을 계산하며 이에 대한 증발률을 적용하였다.

3.2. 정의

정의 1 (원시 온톨로지). 원시 온톨로지(Primitive Ontology)는 사용자가 작성한 초기의 온톨로지이며 PO 로 표기한다. PO 는 여러 개의 RDF 트리플로 구성된다. 만약 사용자가 *subject* s , *predicate* r , *object* o 로 구성된 RDF 트리플을 n 개 작성했다고 할 때 PO 는 이들의 집합을 의미하며 (식 2)와 같이 표현된다.

$$PO = \{(s_1, r_1, o_1), (s_2, r_2, o_2), \dots, (s_i, r_i, o_i), \dots, (s_n, r_n, o_n)\}$$

$$(1 \leq i \leq n) \tag{식 2}$$

정의 2 (매개 온톨로지). 매개 온톨로지(Mediation Ontology) MO 는 m 명의 사용자들이 작성한 원시 온톨로지 PO 들의 집합이다. 이 때 $S = \{PO_1, PO_2, \dots, PO_j, \dots, PO_m\}$ ($1 \leq j \leq m$)이라고 한다면 US 를 다음 (식 3)과 같이 쓸 수 있다.

$$US = \cup_{j=1}^m PO_j \tag{식 3}$$

따라서 다음 (식 4)에 의하여 매개 온톨로지는 US 로 정의될 수 있다.

$$MO = US$$

$$= \cup_{j=1}^m PO_j$$

$$= \{(s, r, o) | (s, r, o) \in PO \text{ for some } PO \in MO\}$$

$$= \{(s_1, r_1, o_1), (s_2, r_2, o_2), \dots, (s_k, r_k, o_k)\}$$

for $k = \sum_{j=1}^m |PO_j|$ (식 4)

정의 3 (대중성 페르몬). 대중성 페르몬(Popularity Pheromone) τ 는 어떤 *subject* s 와 *object* o 에 대해 사용자들이 기입한 관계 r 의 대중성을 평가하는 지표이다. 대중성 페르몬 τ 는 ACO 의 계산 모형을 따른다. 페르몬의 증발률이 0 에서 1 사이의 실수값인 ρ 이라고 하고 k 명의 사용자가 원시 온톨로지를 작성했다고 할 때 대중성 페르몬 τ 는 (식 7)과 같이 계산한다. (식 7)에서 $|A|$ 와 $|B|$ 는 (식 5)와 (식 6)과 같이 정의하는데 어떤 s 와 어떤 o 를 공통적으로 가지고 있는 MO 의 원소 개수를 $|A|$ 라고 표기하며 어떤 s, r, o 를 공통적으로 가지고 있는 MO 의 원소 개수를 $|B|$ 라고 표기한다. 다시 말해, MO 에서 *subject* s 와 *object* o 가 같은 원소들만으로 구성된 집합을 A 라고 정의하며 이 때 A 는 다음 (식 5)와 같이 정의된다.

$$A = \{(s_1, r_1, o_1), (s_2, r_2, o_2), \dots, (s_i, r_i, o_i)\}$$

($0 < i \leq |MO|$) (식 5)

그리고 집합 A 에서 *subject* s , *relation* r , *object* o 가 모두 같은 원소들만으로 구성된 집합을 B 라고 할 때 B 는 다음 (식 6)과 같이 정의된다.

$$B = \{(s_1, r_1, o_1), (s_2, r_2, o_2), \dots, (s_k, r_k, o_k)\}$$

($1 < k \leq |A|$) (식 6)

대중성 페르몬 τ 는 각 (s, r, o) 에 대하여 다음 식 (식 7)과 같이 정의된다.

$$\tau_{so}(1) = (1-\rho) \cdot \tau_{so}(0) + \rho \cdot (|B| / |A|)$$

$$\tau_{so}(k+1) = (1-\rho) \cdot \tau_{so}(k) + \rho \cdot (|B| / |A|) \tag{식 7}$$

Definition 4 (대중 온톨로지). 대중 온톨로지(Popular Ontology) LO 는 MO 에서 τ 값이 사용자가 설정한 임계치 ϵ 을 넘은 (s, r, o) 중에서 τ 값이 가장 높은 것으로만 구성된 온톨로지이다. 여기서 ϵ 는 임계치를 의미하는 사용자 변수이며 0 에서 1 사이의 실수이다. 만

약 τ 의 값이 ϵ 를 초과한다면 대중성이 있는 것으로 판단하고 대중 온톨로지로 출력하며 그렇지 않을 경우 대중 온톨로지로 출력하지 않는다. ϵ 값의 크기가 커지거나 작아질수록 대중 대중온톨로지의 정확성이 달라진다. 대중 온톨로지 LO 는 MO 가 m 개의 원소를 가지고 있으며 $\tau(s,r,o) > \epsilon$ 일 때 다음 (식 8)과 같이 표현된다.

$$LO = \bigcup_{i=1}^m (MAX(\tau(s_i, r_i, o_i))) \quad (m = |MO|, \tau(s,r,o) > \epsilon) \quad (식 8)$$

3.3. ACO 를 이용한 온톨로지의 생성 단계

본 논문에서 제안하는 온톨로지의 생성방법은 다음 리스트와 같이 4 개 단계로 구성되어있다.

- 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

첫번째 단계에서는 사용자들이 원시 온톨로지를 생성하는 단계이다. 이 단계에서 원시온톨로지가 생성되어 온톨로지 저장소로 전송된다. 두번째 단계에서는 전송받은 원시 온톨로지를 취합하고 저장한다. 세번째 단계에서는 원시온톨로지를 이용하여 매개온톨로지를 생성하고 동일한 개념에 대한 빈도를 계산한다. 마지막으로 네번째 단계에서는 대중성 페르몬을 계산하여 대중 온톨로지를 생성한다.

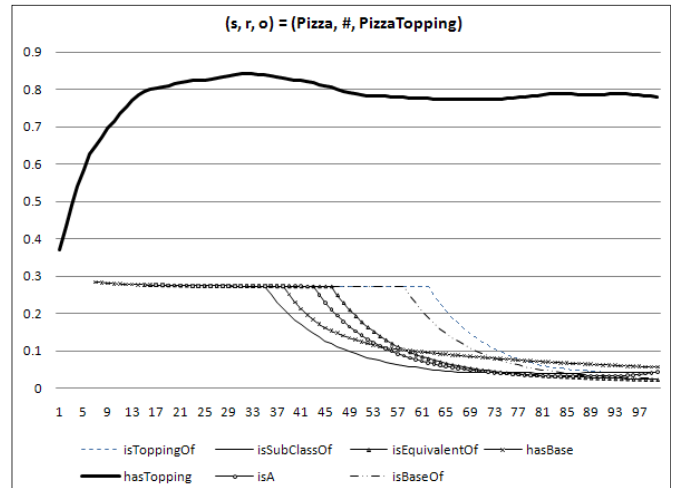
4. 구현 및 실험

본 논문에서 제안한 방법을 검증하기 위하여 우리는 C# 프로그래밍 언어를 이용하여 다수의 사용자가 원시 온톨로지를 생성하여 서버에 전송할 수 있도록 사용자 프로그램을 구현하였으며 서버에서 원시 온톨로지를 입력 받아 매개 온톨로지를 생성하고 ACO 를 통해 페르몬을 계산하여 대중 온톨로지를 생성하는 서버 프로그램을 작성하였다.

본 논문에서는 제안하는 방법의 검증을 위하여 많은 연구에서 예제 온톨로지 사용하고 있는 맨체스터 대학교의 피자 온톨로지[20]를 비교 대상으로 활용하였다. 우리는 100 명의 온톨로지 및 시맨틱 웹의 비전문가에게 피자 및 피자 토핑에 대한 키워드를 순서 없이 제시하였으며 구현한 사용자 프로그램을 활용하여 피자와 피자 토핑에 대한 원시 온톨로지를 작

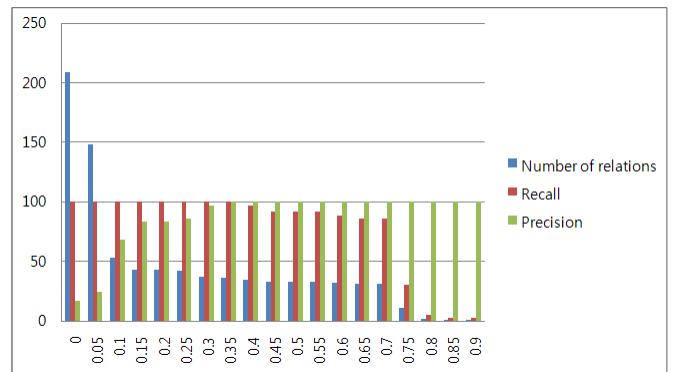
성하여 서버로 전송하도록 하였다. 따라서 실험 및 검증에 쓰인 원시 온톨로지는 총 100 개이며 이 100 개의 원시 온톨로지를 취합하여 209 개의 연결관계를 가진 매개 온톨로지를 생성하였다. 그리고 대중 온톨로지를 생성할 때 입력한 초기값으로는 증발률 ρ 는 0.1, 페르몬 τ 의 초기값은 0.3 으로 설정하였다.

다음 (그림 1)은 100 명의 사용자에게 입력받은 209 개의 연결 관계 중에서 두 클래스, 'Pizza'와 'PizzaTopping'의 관계에 대해 사용자들이 입력받은 순서대로 페르몬 값의 변화에 대한 그래프이다.



(그림 1) 페르몬 업데이트

(그림 1)의 가로축은 원시 온톨로지의 입력 순서이며 세로축은 페르몬 값을 의미한다. 보이는 바와 같이 원시 온톨로지를 입력받은 순서에 따라 페르몬 값이 변화하는 것이 보이는데 (Pizza, isToppingOf, PizzaTopping)을 입력한 사용자가 많을수록 그에 해당하는 페르몬 값이 점진적으로 증가하며 다른 관계를 입력한 사용자는 상대적으로 적으므로 초기값 0.3 에서 점진적으로 하락하는 것을 볼 수 있다. 다시 말해 어떤 s 와 o 에 대해 사용자들이 일반적으로 알고 있는 관계 r 이 있다면 그 (s, r, o) 는 초기값에서 점진적으로 증가하여 1 에 가까운 값으로 수렴하며 그렇지 않은 경우에는 0 에 가까운 값으로 수렴하게 된다.



(그림 2) 임계치의 변화에 따른 정확도와 재현률

위 (그림 2) 는 임계치 ϵ 의 변화에 따른 대중 온톨

로지의 정확도와 재현률을 보인다. 가로축은 임계치가 0 일 때부터 0.9 일 때까지를 보이며 세로 축은 대중 온톨로지로 출력된 관계의 숫자와 정확도, 재현률의 수치를 보인다. (그림 2)에서 보이는 바와 같이 임계치 ε 가 커지면 커질수록 출력되는 대중 온톨로지의 크기가 작아짐을 알 수 있으며 정확도와 재현률의 결과값이 좋아짐을 알 수 있다. 반대로 임계치 ε 가 작아질 수록 대중 온톨로지의 크기가 커지고 정확도와 재현률의 결과가 나빠지게 된다. 따라서 임계치 ε 는 0 이나 1 처럼 극단적인 값을 설정했을 때 적절한 대중 온톨로지를 기대하기 어려우며 0.35 에서 0.7 사이의 값으로 설정하였을 때 적절한 대중 온톨로지의 크기를 유지하면서 좋은 정확도와 재현률을 기대할 수 있다.

5. 결론

웹 2.0 의 대두를 통해 최근의 인터넷 환경은 분산 컴퓨팅 및 서비스 기반 인터넷으로 빠르게 변화하고 있다. 이 가운데, 웹 2.0 으로의 변화를 가속화 시키는 중요 기술로서 집단지성이 다양한 분야에서 활용이 되고 있다. 집단지성은 분산 환경하에서 사용자들의 지식 및 경험이 축적되어 대중적으로 활용 가능한 형태의 정보를 만들어 낼 수 있다는 것이 그 특징이다.

한편 2001 년 시맨틱 웹이 제창된 이후 시맨틱 웹의 중요한 기술 요소인 온톨로지의 생성은 다양한 연구와 개발이 이루어져 왔음에도 불구하고 현재까지 대량의 온톨로지를 생성할 수 있는 방법은 제시되지 못하고 있다.

본 논문에서는 웹 2.0 의 중요 개념 중에 한가지인 집단지성을 활용하여 시맨틱 웹 분야의 오랜 난제였던 온톨로지의 대량 생성을 가능케 하는 시스템을 제안하였으며 ACO 를 이용한 협업적 온톨로지의 생성 방법을 제시하였다. 본 논문에서 제시한 온톨로지의 생성 시스템 및 방법은 웹 2.0 기반 사용자 어플리케이션에서 원시 온톨로지를 제작하고 제작된 원시 온톨로지는 온톨로지로서 전송되어 매개 온톨로지로서 통합될 수 있는 구조를 가지고 있다. 따라서 본 논문에서 제안하는 방법을 활용하면 다양한 도메인의 온톨로지를 비록 전문가가 아니지만 다수의 사용자가 참여하여 대량으로 생성할 수 있다. 그리고 본 논문에서 제안한 방법으로 생성된 온톨로지들은 시맨틱 웹 기반 지능형 서비스의 연구 및 개발에 있어 중요한 디딤돌이 될 것으로 기대한다.

참고문헌

[1]Mark Greaves, "Semantic Web and Web 2.0" Web Semantics: Science, Services and Agents on the World Wide Web 6 (2008) 1-3
 [2]Luke Tredinnick, "Web 2.0 and Business : A pointer to the intranets of the future?", Business Information Review, 2006
 [3]T.Berners-Lee, J.Hendler, and O.Lasilla "The Semantic Web", Scientific American, May 2001.

[4]Davies, John, Fensel, Dieter, Harmelen Frank Van. 2003. Towards the semantic web: ontology-driven knowledge management. [s. l.]: Wiley.
 [5]Dean Jones, Trevor Bench-Capon, Pepijin Visser. 1998. Methodologies For Ontology Development. In Proc.
 [6]Protege, <http://protege.stanford.edu>
 [7]Protege OWL API Programmer's Guide, <http://protege.stanford.edu/plugins/owl/api/guide.html>
 [8]Koji Zettus & Yasushi Kiyoki, "Towards Knowledge Mangement Based on Harnessing Collective Intelligence on the Web", LNAI 4248, 2006
 [9]James Surowiecki, "Wisdom Of Crowds", Random House, 2005
 [10]OWL : <http://w3g.org/TR/2004/REC-owl-features-20040210/>
 [11]Rubén Prieto Díaz. A Faceted Approach to Building Ontologies. the 2003 IEEE International Conference on Information Reuse and Integration, IRI - 2003, Las Vegas, NV, USA, October 27-29, 2003.
 [12]Vijayan, Sugumaran, Veda C. Storey. Ontologies for conceptual modeling: their creation, use, and management. Data & Knowledge Engineering Volume 42 , Issue 3, September 2002, pp. 251 - 271
 [13]Thanh Tho Quan, Siu Cheung Hui, Alvis Cheuk M. Fong, Tru Hoang Cao. Automatic Generation of Ontology for Scholarly Semantic Web. ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, pp 726-740, November 7-11, 2004.
 [14]J.M. Cueva Lovelle et al. Swarm Intelligent Surfing in the Web" ICWE 2003, LNCS 2722, pp. 431-440, 2003
 [15]James Surowiecki, "Wisdom Of Crowds", Random House, 2005
 [16]Szuba T.: Computational Collective Intelligence. Wiley and Sons NY. 2001, Wiley Book Series on Parallel and Distributed Computing.
 [17]RSS Advisory Board. Really Simple Syndication: RSS 2.0.1 Specification (revision 6). Retrieved March 2, 2006, <http://www.rssboard.org/rss-2-0-1-rv-6>.
 [18]Dorigo, M., Blum, C.: Ant colony optimization theory: A survey. Theor. Comput. Sci. **344**, 243-278 (2005)
 [19]Siorpaes, K. and M. Hepp, myOntology: The Marriage of Collective Intelligence and Ontology Engineering, in Proceedings of the Workshop Bridging the Gap between Semantic Web and Web 2.0 at the ESWC 2007. 2007, Innsbruck, Austria
 [20] <http://www.co-ode.org/ontologies/pizza/2007/02/12/>