

잠재적인 참조를 고려한 논문 랭킹 방안

황 세 미[†] · 배 덕 호^{**} · 김 상 욱^{***}

요 약

본 논문에서는 예전 논문의 기득권 현상을 해결하기 위해 한 논문이 현재 다른 논문들로부터 받은 참조뿐만 아니라, 해당 논문의 잠재적인 참조도 함께 고려하는 랭킹 방안을 제안한다. 더 나아가, 논문의 정확한 랭킹 측정을 위해 두 논문 간의 내용 연관성도 함께 고려하는 랭킹 방안을 제안한다. 마지막으로, 실제 논문 데이터를 이용한 다양한 실험들을 통해, 제안 방안의 우수성을 입증한다.

키워드 : 논문 랭킹, 잠재적인 참조, RWR, WCP

Scientific Literature Ranking Considering Latent Citations

Se-Mi Hwang[†] · Duck-Ho Bae^{**} · Sang-Wook Kim^{***}

ABSTRACT

In this paper, to solve a vested interests of old papers in scientific literature ranking, we propose novel method that considers not only the current citations from other published papers but also the latent citations of papers to be published in the future. Furthermore, the method also considers the relevance of contents in the citing and cited papers. Finally, we verify the superiority of our proposed method through extensive experiments.

Keywords : Literature Scientific Ranking, Latent Citation, RWR, WCP

1. 서 론

최근, 다양한 검색 사이트들의 등장으로 인해, 사용자는 간단한 질의어 입력을 통해 논문을 검색할 수 있게 되었다. 그러나 검색 사이트는 해당 질의어에 대한 매우 많은 수의 결과를 보여주므로, 사용자가 이를 일일이 검토하여 실제로 원하는 논문을 발견하기까지는 많은 노력이 필요하다. 따라서 검색 사이트는 사용자의 편의를 위해 사용자가 원할 것으로 예상되는 논문을 질의 결과의 상위에 위치시켜줄 필요가 있다.

기존에는 주로 웹 페이지를 대상으로 하는 랭킹 방안들이 연구되어 왔었다. 널리 알려진 웹 페이지 랭킹 방안으로는

PageRank[1], HITS[2] 등이 존재한다. 이들은 웹 연결망을 웹 페이지를 노드로, 웹 페이지 간의 하이퍼링크를 방향성 에지로 하는 그래프로 모델링한 후, 해당 그래프를 분석하여 각 웹 페이지의 권위 (authority)를 측정하는 방안이다. 이 때, 웹 페이지의 권위가 높을수록 해당 웹 페이지는 사용자가 원하는 검색 결과일 것으로 간주한다.

논문에는 웹 페이지의 하이퍼링크와 대응되는 참조 정보가 존재한다. 그러나 웹 페이지와 달리 논문은 발행된 이후에는 참조 정보를 수정할 수 없으므로, 자신보다 먼저 발행된 논문만 참조할 수 있는 특징을 가진다. 이로 인해, 최근에 발행된 논문은 예전에 발행된 논문에 비해 참조 받을 기회가 부족하여 상대적으로 받은 참조의 수가 적게 된다. 따라서 논문을 위한 랭킹 방안은 논문의 권위를 측정하는데 이러한 논문의 참조 특성을 반영하여야 한다.

기존의 논문 랭킹 방안에는 Browsing based method[3], CiteRank[4], PopRank[5] 등이 존재하였다. 기존 방안들은 논문을 노드로, 참조 관계를 방향성 에지로 하는 그래프로 모델링한 후, 각 논문의 권위를 측정한다. 그러나 대부분의 기존 논문 랭킹 방안들은 위에서 언급한 논문의 참조 특성

* 본 연구는 지식경제부 및 정보통신산업진흥원의 IT융합 고급인력과정 지원 사업(NIPA-2011-C6150-1101-0001)의 지원과 2012년도 두뇌한국 21 사업의 지원과 NHN(주)의 지원을 받았습니다. 그러나 본 논문에서 제시된 의견이나 결론, 또는 권고 등은 온전히 저자(들)의 것이며, 반드시 지원회사의 입장을 대변하는 것은 아닙니다.

† 준 회원 : 한양대학교 전자컴퓨터통신공학과 석사

** 준 회원 : 한양대학교 전자컴퓨터통신공학과 박사과정

*** 종신회원 : 한양대학교 정보통신대학 정보통신학부 교수

논문접수 : 2011년 10월 5일

수정일 : 1차 2011년 11월 22일

심사완료 : 2011년 11월 24일

을 고려하지 않아, 최근에 발행된 품질이 좋은 논문일지라도 예전 논문에 비해 상대적으로 낮은 권위 점수를 부여받게 되는 문제가 발생하게 된다.

이렇듯 기존의 논문 랭킹 방안들은 예전부터 권위를 인정 받은 논문들을 찾는 것에는 유용하지만, 최근에 발행된 품질이 좋은 논문들을 찾기에는 적합하지 않다. 그러나 논문 검색 사이트의 주 이용자인 연구자들은 예전의 기반 연구뿐만 아니라 최근의 연구 동향도 중요시 여기며, 논문 작성 시 최신 논문을 인용하고자 하는 경향을 보인다. 따라서 품질이 좋은 최신 논문들을 검색 결과의 상위에 올려줄 필요가 있다.

본 논문에서는 현재 해당 논문이 다른 논문들로부터 받은 권위뿐만 아니라, 해당 논문이 미래에 받을 잠재적인 권위도 함께 고려하는 랭킹 방안을 제안한다. 이를 통해, 예전 논문과 최신 논문이 공정한 경쟁을 할 수 있게 한다. 더 나아가, 좀 더 정확한 논문의 권위 측정을 위해 두 논문간의 내용 연관성을 고려하는 랭킹 방안을 제안한다. 마지막으로, 다양한 실험을 통해 제안하는 방안의 정확성 및 효용성을 검증한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 기존의 논문 랭킹 방법들에 대해 설명한다. 제 3장에서는 본 논문에서 제안하는 방안을 상세하게 설명한다. 마지막으로, 4장에서는 실제 논문 데이터를 이용한 다양한 실험을 통해, 제안 방안의 우수성을 보인다.

2. 관련 연구

논문 랭킹 방안은 논문 분야 연구의 중요한 이슈중의 하나이다. 기존 논문 랭킹 방안으로는 Browsing-based model[3], PopRank[5], CiteRank[4] 등이 존재한다.

Browsing-Based model (BBM)[3]은 논문의 권위를 계산하기 위해 논문들 간의 참조 관계뿐만 아니라 논문을 작성한 저자와의 저작 관계까지 고려한 방법이다. 논문을 노드로, 논문간의 참조 관계를 방향성 에지로 각각 간주하고, 동일한 저자가 작성한 논문들 사이에 양방향성 에지를 추가로 연결한 그래프로 모델링 한다. 이는 권위 있는 논문을 작성한 저자의 다른 논문들에게 더 많은 권위 점수를 부여하기 위해서이다. BBM은 이렇게 모델링한 그래프에서 random walk with restart (RWR)[1]을 수행하여 각 논문의 권위 점수를 계산한다. 그러나 BBM은 논문의 최신성을 전혀 고려하지 않으므로, 예전 논문들이나 해당 분야의 대가가 쓴 논문들이 상대적으로 높은 권위 점수를 갖는 문제점이 발생한다.

PopRank[5]는 논문의 권위를 계산하기 위해 논문들 간의 참조 관계뿐만 아니라 논문을 작성한 저자와의 저작 관계, 해당 논문이 발행된 학회, 그리고 web popularity까지 고려한 방법이다. 논문을 노드로, 논문 간의 참조 관계를 방향성 에지로 각각 간주하고, 동일한 저자가 작성한 논문들과 동일한 학회에서 출판된 논문들, 그리고 동일한 학회에 논문

을 게재한 저자들 사이에 양방향성 에지를 연결한 그래프로 모델링 한다. 이는 권위 있는 논문을 작성한 저자의 다른 논문들과 권위 있는 학회에 게재된 다른 논문들에게 더 많은 권위 점수를 부여하기 위해서이다. PopRank는 해당 그래프에서 RWR을 수행할 뿐만 아니라 논문, 저자, 학회의 Web popularity 점수를 계산한 후, 이 둘을 합하여 논문의 최종 권위 점수를 계산한다. 그러나 PopRank는 BBM과 마찬가지로 논문의 최신성을 전혀 고려하지 않으므로, 예전 논문들이나 해당 분야의 대가가 쓴 논문들, 그리고 명성이 높은 학회에 실린 논문들이 상대적으로 높은 권위 점수를 갖는 문제점이 발생한다.

CiteRank[4]는 권위 있는 논문뿐만 아니라 최신의 논문 또한 검색 결과의 상위에 랭크 시키고자 시도한다. 이를 위해, 논문들 간의 참조 관계뿐만 아니라 논문의 출판 연도도 함께 고려한다. CiteRank는 논문을 노드로 하고 논문들 간의 참조 관계를 방향성 에지로 하여 그래프를 모델링한 후, RWR을 수행하여 각 논문의 권위 점수를 계산한다. CiteRank는 최신 논문에 높은 랭크를 부여하기 위하여 최근에 발행된 논문의 최초 점수를 높게 부여하는 방법을 사용한다. 그러나 CiteRank는 발행 연도가 동일하면 동일한 초기 값을 부여함으로써, 품질이 좋은 최신 논문에 높은 권위 점수를 효율적으로 부여하지 못하는 문제점이 존재한다.

3. 제안 방안

본 논문에서는 제안하는 랭킹 방안의 목표는 다음과 같다.

목표 1: 다른 논문들로부터 많은 참조를 받거나 혹은 많은 참조를 받은 논문들로부터 참조를 받은 논문을 상위에 랭크 시키고자 한다. 만약 논문 p 가 많은 논문들로부터 참조를 받거나 많은 참조를 받은 논문들로부터 참조를 받았다면, 논문 p 는 품질이 좋고 해당 분야에서의 영향력이 크다는 것을 의미한다.

목표 2: 최근에 발행된 좋은 논문을 상위에 랭크 시키고자 한다. 최신 논문들은 이후에 발행된 논문들의 수가 예전 논문에 비해 상대적으로 적으므로 예전 논문에 비해 참조 받은 기회가 부족하다. 그럼에도 불구하고, 만약 최신 논문 p 가 발행 연도 대비 일정 이상의 참조를 받았다면, 최신 논문 p 는 미래에 많은 참조를 받을 우수한 논문이라는 것을 예상할 수 있다.

목표 3: 랭킹을 결정하는데 있어 논문들 간의 내용 연관성을 고려하고자 한다. 논문 p 이후에 발행된 유사한 논문들은 논문 p 와의 차이점을 설명하기 위해 논문 p 를 참조할 의무를 지닌다. 만약, 논문 p 를 참조하지 않았다면, 논문 p 의 품질이 낮거나 해당 분야에서의 중요도가 높지 않다는 것을 간접적으로 의미한다. 따라서 유사한 논문들로부터 참조를 많이 받거나, 이후에 참조 받을 가능성이 높은 논문이 해당 분야에서 영향력이 크거나 미래에 주목 받을 논문이라고 볼 수 있다.

본 논문에서는 해석 가능한 결과 도출을 위해 확률 모델 (probabilistic framework)을 채택한다. 이를 위해 웹 페이지, 블로그, 논문 랭킹 등에서 널리 쓰이고 있는 RWR[1]을 채택한다. RWR을 수행하기 위해서는 논문과 논문들 간의 참조 관계를 그래프 형태로 모델링하는 과정이 필요하다. 모델링 시, 한 편의 논문은 하나의 노드로, 논문들 간의 참조는 방향성 에지로 간주한다.

수식 1은 RWR의 개념을 나타낸다.

$$\vec{r} = (1 - c)W\vec{r} + c\vec{e} \quad (1)$$

- \vec{r} : $n \times 1$ 랭킹 결과 벡터
- \vec{e} : $n \times 1$ 리스트 벡터¹⁾
- c : 리스트 확률
- W : $n \times n$ 정규화된 인접 행렬

RWR은 참조 관계 그래프에서 논문들의 권위 점수를 효과적으로 측정할 수 있다. 연구자가 임의의 논문 한 편을 선택하여 읽기 시작한다고 가정하자. 연구자는 $1-c$ 의 확률로 해당 논문이 참조한 논문들을 읽거나 혹은 c 의 확률로 리스트 벡터 \vec{e} 에 의해 결정된 임의의 논문들을 읽는다. 이러한 단계를 반복하면, 참조 관계 그래프의 모든 논문들이 연구자에 의해 읽혀질 수렴된 확률 분포 \vec{r} 을 구할 수 있다. 이 때, r_i 는 연구자가 논문 i 를 읽을 확률을 나타낸다. r_i 가 크다는 것은 논문 i 의 품질이 좋고 이로 인해 다수의 우수한 논문들로부터 참조 받았다는 것을 의미한다. 따라서 RWR은 목표 1을 만족한다.

목표 2를 만족하기 위해, 한 논문이 현재까지 받은 참조 뿐만 아니라, 미래에 추가적으로 받을 잠재적인 참조도 함께 고려한다. 이를 위해 한 논문이 미래에 참조 받을 잠재력을 나타내는 Citation Probability (CP)를 제안한다. $CP(p)$

$$CP(p) = \frac{\text{count}(C(p))}{\text{count}(T(p))} \quad (2)$$

- $CP(p)$: 논문 p 의 CP
- $\text{count}(A)$: 집합 A 에 속하는 논문의 수
- $C(p)$: 논문 p 를 참조하는 논문들의 집합
- $T(p)$: 논문 p 이후에 발행된 동일한 분야의 모든 논문들의 집합

1) 벡터의 값은 각 노드에서 리스트 할 확률을 의미하며, 그래프의 전체 노드 수가 n 개일 때, 벡터 e 의 각 인자 e_i 는 $1/n$ 의 균일한 값으로 설정한다.

는 논문 p 이후에 발행된 동일한 분야의 논문들이 논문 p 를 참조한 비율을 의미한다. 이는 해당 논문이 발행된 이후의 동향을 의미하므로, 결국 미래에 발행될 동일한 분야의 논문들이 해당 논문을 참조할 확률을 의미한다.

논문 p 의 $CP(p)$ 는 $C(p)$ 가 클수록, $T(p)$ 는 작을수록 커진다. 예전 논문의 경우, 최신 논문에 비해 상대적으로 참조 받을 기회가 많으므로 $C(p)$ 가 커진다. 반면, 논문 p 이후에 많은 논문들이 발행되었으므로, $T(p)$ 도 커진다. 이와 반대로, 최신 논문의 경우 $C(p)$ 는 작아지지만, $T(p)$ 역시도 작아진다. 이렇듯, 절대적인 참조 받은 수가 아닌 참조 받은 비율을 고려함으로써, 예전 논문과 최신 논문이 동등한 경쟁이 가능하게 된다.

RWR은 현재 시점에서의 해당 논문의 권위 점수를 의미하는데 반해, 목표 2를 위한 CP는 해당 논문이 미래에 참조 받을 가능성을 의미한다. 따라서 본 논문에서는 두 점수를 모두 고려하여 논문의 최종 권위 점수를 계산한다. 최종 권위 점수를 계산하는 방안은 뒤에서 언급한다.

마지막으로 목표 3을 만족하기 위해, RWR과 CP에 각각 논문들 간의 내용 연관성을 반영한다. 첫째, RWR 수행 시, 유사한 논문으로부터 참조를 받은 논문이 높은 점수를 받도록 가중치를 부여한 인접행렬을 이용한다.

$$w_{i,j} = \text{sim}(i,j) / \sum_{k \in \text{CitedBy}(i)} \text{sim}(i,k) \quad (3)$$

- $w_{i,j}$: 인접 행렬의 i 행 j 열의 원소
- $\text{sim}(a,b)$: 두 논문간의 유사도²⁾
- $\text{CitedBy}(i)$: 논문 i 를 참조한 논문 집합

이를 통해, 논문들은 자신이 참조한 논문들 중 연관성이 높은 논문들에게 더 높은 확률로 랜덤 워크를 수행할 수 있게 된다. 따라서 연관성이 높은 논문들로부터 참조를 많이 받은 논문이 그렇지 않은 논문보다 더 큰 권위 점수를 가지게 되며, 이는 목표 3을 만족한다.

둘째, 논문의 미래 잠재력을 계산할 때, 모든 참조를 동일하게 고려하지 않고, 연관성이 높은 논문으로부터의 참조를 더 중요시 여긴다. 수식 4는 CP에 내용 연관성을 반영한 Weighted Citation Probability (WCP)를 나타낸다.

앞에서도 설명하였듯이, 논문 p 와 유사한 논문들은 논문 p 를 참조할 의무를 지닌다. 반대로, 논문 p 와 유사하지 않은 논문들은 논문 p 를 참조할 필요가 없다. 따라서 논문들 간의 내용 유사도를 고려함으로써, 해당 논문이 미래의 잠재력을 좀 더 정확하게 측정할 수 있다.

2) 본 논문에서는 두 논문간의 유사도 계산을 위해 다음의 논문을 참고하였다 [9]. 한 가지 언급할 것은 두 논문간의 유사도 계산을 위해 다양한 방안들을 사용 가능하며, 유사도 계산 방안에 대한 특별한 제약을 두지 않는다.

$$WCP_{(p)} = \frac{\sum_{c \in C_{(p)}} sim(p, c)}{\sum_{t \in T_{(p)}} sim(p, t)} \quad (4)$$

- $WCP_{(p)}$: 논문 p 의 WCP 점수
- $C_{(p)}$: 논문 p 를 참조하는 논문들의 집합
- $T_{(p)}$: 논문 p 이후에 발행된 동일한 분야의 모든 논문들의 집합

제안하는 랭킹 방안의 최종 권위 점수는 RWR과 WCP의 두 확률의 곱으로써 계산한다. 이를 통해, 현재까지 축적된 논문의 권위뿐만 아니라 미래에 참조 받을 가능성까지 고려한 논문의 권위 점수를 구할 수 있다.

$$\text{최종 권위 점수}_{(p)} = RWR_{(p)} \times WCP_{(p)} \quad (5)$$

- $RWR_{(p)}$: 논문 p 의 수렴 단계의 RWR 점수

4. 실험

4.1 실험 환경

실험에는 2009년 3월에 다운로드한 DBLP 데이터[6]를 사용한다. DBLP 데이터에는 참조 정보가 포함되어 있지 않으므로, Libra 데이터[3]를 통해 논문들 간의 참조 정보를 수집하였다.

대표적인 데이터 마이닝 기법인 클러스터링, 연관 규칙, 그래프 마이닝의 총 3개 분야를 대상으로 실험을 수행한다. <표 1>은 각 분야의 기본 통계를 보인다.

<표 1> 각 분야의 통계

	정답 셋 수	논문 수	참조 수
Clustering	37	1,464	6,892
Association rules	60	1,280	8,262
Graph mining	21	334	1,389

정답 셋 작성을 위해 널리 알려진 데이터 마이닝 책[7]의 각 장의 참고 문헌에 기술된 논문들을 정답 셋으로 간주하였다. 단, 참고 문헌 [8]에서 논문은 발행 후 3년이 지나야

충분한 참조를 받는 것으로 나타났으므로, 발행된 지 3년 미만인 논문은 정답 셋에서 제외하였다. 각 분야 별로 정답 셋에 속한 논문들과 참조 관계에 있는 논문들은 동일한 분야에 속하는 논문들로 간주하였다.

4.2 실험 결과

본 절에서는 제안하는 방안의 우수성을 보이기 위해 기존의 논문 랭킹 방안인 RWR, CiteRank, BBM, PopRank와 제안 방안의 성능을 비교한다. 이를 위해, 리스타트 확률 c 를 널리 통용된 구간인 0.05에서부터 0.5까지 0.05단위로 변화시키며 각 방안들의 mean average precision (MAP)을 측정하였다. 이 때, 제안 방안과 기존 방안의 매개 변수 c 의 설정을 위해 3-fold cross validation을 수행하였다. Cross validation 결과, 제안 방안과 기존 방안에서 c 가 0.5일 때, 가장 좋은 성능을 보였다. 따라서 이후의 모든 실험에서 기존 방안 및 제안 방안의 c 는 모두 0.5로 설정하였다.

<표 2>는 총 5가지 방법들의 MAP을 측정한 것이다. 실험 결과, 제안 방안의 경우, 기존의 다른 방안에 비해 우수한 결과를 보임을 알 수 있다. 이는 제안하는 논문 랭킹 방안은 현재 해당 논문이 다른 논문들로부터 받은 권위뿐만 아니라, 해당 논문의 미래의 잠재력을 함께 고려하여 최신의 품질 좋은 논문들도 상위에 랭크시킴으로써, 높은 MAP을 보이는 것을 알 수 있다. 이러한 경향은 <표 3>에서 더욱 잘 나타난다.

흥미롭게도, BBM의 경우, RWR보다 낮은 MAP을 보였다. 이는 BBM은 저자 및 학회 정보를 사용하기 때문인 것으로 추측된다. 이로 인해, 유명한 저자가 작성한, 혹은 좋은 학회에 실린 논문의 경우, 참조를 거의 받지 못하더라도 상위에 랭크 시키게 되어, MAP이 낮아졌다.

<표 2> MAP 결과

	RWR	Cite Rank	BBM	Pop Rank	제안 방안
MAP	0.589	0.662	0.558	0.627	0.687

각 방법들의 최신 논문의 추출 여부를 확인하기 위해, 정답 셋에 속하는 논문들의 평균 연도를 기준으로 정답 셋을 예전 논문과 최신 논문으로 분류한 후, 각각에 대해 top 30에서의 평균 recall을 측정하였다.

<표 3>은 그 결과를 나타낸다. 실험 결과, 제안 방안의 경우 예전 논문은 물론 최신 논문에 대한 recall이 다른 방안들에 비해 모두 높다는 것을 알 수 있다. RWR, BBM, PopRank의 경우, 최신 논문의 점수를 올려주는 방안이 존재하지 않으므로, 최신 논문에 대한 recall이 낮게 나타남을 알 수 있다. CiteRank의 경우, 최신 논문의 권위 점수를 올

3) <http://libra.msra.cn>

〈표 3〉 top 30의 recall 결과

	RWR	Cite Rank	BBM	Pop Rank	제안 방안
Old	0.726	0.717	0.635	0.660	0.742
Recent	0.232	0.389	0.293	0.324	0.450
Average	0.479	0.553	0.464	0.492	0.596

려주는 방안이 존재하여 RWR, BBM, PopRank에 비해 최신 논문에 대한 recall이 높지만, 제안 방안에 비해서는 많이 낮은 것을 알 수 있다.

5. 결 론

본 논문에서는 논문의 최신성과 연관성을 반영한 논문 랭킹 방안을 제안하였다. 논문의 공헌은 다음과 같다. 첫째, 논문의 최신성을 고려하기 위해, 해당 논문이 다른 논문들로부터 받은 권위뿐만 아니라, 해당 논문의 미래의 잠재력도 함께 고려하는 랭킹 방안을 제안하였다. 둘째, 두 논문간의 내용 연관성을 고려하여 논문의 랭킹의 정확성을 높였다. 마지막으로, 실제 논문 데이터를 이용한 다양한 실험들을 통해, 제안 방안의 우수성을 입증하였다.

참 고 문 헌

[1] L. Page et al., The PageRank Citation Ranking: Bringing Order to the Web, *Technical Report*, Stanford University, 1999.

[2] J. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM*, Vol.46, No.5, pp.604-632, 1999.

[3] S. Yan and D. Lee, "Toward Alternative Measures for Ranking Venues: a Case of Database Research Community," In Proc. *Int'l Conf. on Digital Libraries*, JCDL, pp.235-244, 2007.

[4] D. Walker et al., "Ranking Scientific Publications using a Simple Model of Network Traffic," *Journal of Statistical Mechanics: Theory and Experiment*, 2007.

[5] Z. Nie et al., "Object-Level Ranking: Bringing Order to Web Objects," In Proc. *Int'l Conf. of World Wide Web*, WWW, pp.567-574, 2005.

[6] M. Ley, "DBLP: Some Lessons Learned," In Proc. *Int'l Conf. on Very Large Data Bases*, VLDB, pp.1493-1500, 2009.

[7] J. Han and M. Kamber, *Data Mining: Concepts and Techniques(2nd Edition)*, Morgan Kaufmann, 2006.

[8] E. Rahm and A. Thor, "Citation Analysis of Database Publications," *ACM SIGMOD Record*, Vol.34, No.4, pp.53-59, 2005.

[9] Yoon, S. Kim, and S. Park, "A Link-based Similarity Measure for Scientific Literature," In WWW, pp.1213-1214, 2010.



황 세 미

e-mail : semiya@hanyang.ac.kr
 2009년 2월 서울여자대학교 멀티미디어학과 (학사)
 2011년 2월 한양대학교 전자컴퓨터통신 공학과(석사)
 2011년 3월~현 재 우리파이낸셜 전산팀 재직

관심분야: 사회연결망 분석, 논문 데이터 분석, 데이터 마이닝



배 덕 호

e-mail : dhbae@agape.hanyang.ac.kr
 2006년 2월 한양대학교 컴퓨터공학과(학사)
 2008년 2월 한양대학교 전자컴퓨터통신 공학과(석사)
 2008년 3월~현 재 한양대학교 전자컴퓨터 통신공학과 박사과정

2010년 3월~4월 미국 Carnegie Mellon University, Computer Science Department, Visiting Scholar

관심분야: 임베디드 데이터베이스, 플래시메모리 데이터베이스, 사회연결망 분석, 데이터 마이닝



김 상 욱

e-mail : wook@hanyang.ac.kr

1989년 2월 서울대학교 컴퓨터공학과(학사)

1991년 2월 한국과학기술원 전산학과(석사)

1994년 2월 한국과학기술원 전산학과(박사)

1991년 7월~8월 미국 Stanford University,

Computer Science Department

방문 연구원

1994년 2월~1995년 2월 KAIST 정보전자연구소 전문연구원

1999년 8월~2000년 8월 미국 IBM T.J. Watson Research
Center Post-Doc.

1995년 3월~2000년 8월 강원대학교 컴퓨터정보통신공학부
부교수

2003년 3월~현 재 한양대학교 정보통신대학 정보통신학부 교수

2009년 1월~2010년 미국 Carnegie Mellon University,
Computer Science Department, Visiting Professor

관심분야: 데이터베이스 시스템, 저장 시스템, 트랜잭션 관리,
데이터 마이닝, 멀티미디어 정보 검색, 공간 데이터베
이스/GIS, 주기억장치 데이터베이스, 이동 객체 데이
터베이스/텔레매틱스, 사회 연결망 분석, 웹 데이터
분석