

무선인터넷 서비스를 위한 멀티미디어 통계사서함 구축에 관한 연구

(A Study on Construction of Multimedia Statistic Post Office Box for Wireless Internet Services)

이 중 득(Chong-Deuk Lee)¹⁾ 김 대 경(Dae Kyung Kim)²⁾

요 약

많은 정보들이 디지털 형태로 처리되고 저장됨에 따라 무선 인터넷상에서 멀티미디어 정보들을 서비스하기 위한 새로운 기법들이 개발되고 있다. 본 논문에서는 통계 데이터들간의 유사도를 바탕으로 서로 관련 있는 데이터들을 그룹화 하여 문서들을 주제별로 서비스하기 위한 멀티미디어 통계 사서함(MSPOB)구조를 제안한다. 제안된 구조는 카운트 인덱스구조와 역파일구조를 기반으로 하여 데이터들의 관계성을 결정하게 되며, 객체들간의 의미적 유사도에 따라 제안된 구조가 결정되게 된다.

ABSTRACT

As more and more information is processed and stored in the digital form, many techniques and systems have been developed for service multimedia informations in wireless internet. In this paper, we propose MSPOB(Multimedia Statistics Post Office Box) structure for service datum which are related with similarity to subject a set of documents through grouping. The proposed structure is determined by relationship of datum based on count index and inverted file and is determined it through the semantic similarity between objects

1) 정회원 : 익산대학 정보통신과 교수

논문접수 : 2004. 1. 2.

2) 정회원 : 전북대학교 통계정보학부 교수

심사완료 : 2004. 1. 8.

본 논문은 정보통신부 정보통신 연구진흥원에서 지원하고 있는 정보통신기초기술 연구지원 사업의 연구결과입니다.

1. 서론

최근에 인터넷은 유선인터넷에서 무선인터넷으로 인터넷패러다임이 빠르게 이동하고 있다. 특히 무선 통신 인프라 개선에 따라 무선인터넷 사용인구는 계속 증가할 것으로 기대되며, 무선 통신은 유선통신에 비해 단말사용자가 공간상 제약을 받지 않고 언제 어디서나 누구에게나 통신서비스를 자유롭게 사용할 수 있다는 장점으로 인하여 빠르게 성장하고 있다[1, 4, 7].

그러나 수많은 사용자를 확보하고 있는 무선 통신 서비스는 무선 통신망을 통한 웹 정보의 접근에 관한 연구 및 서비스개발이 미비한 실정이며, 이중 통계 데이터 서비스와 같은 특정 데이터 서비스에 대한 연구는 미흡한 실정이다[7, 8, 9, 10, 11]. 일반적으로 무선 인터넷 서비스는 전송 계층 프로토콜에 대한 연구가 활발히 진행되고 있으며, 사용자 관점에서의 서비스를 지원하기 위해서는 무선 링크 상에서 발생하는 패킷 손실률을 극복해야하며, 전송 과정에서 발생하는 오류율을 최소화해야 한다. 또한 망 자원이 가변적으로 변화되는 변화율을 최소화해야 하며, 제한된 대역폭과 지연시간으로 인한 손실 문제를 고려해야 한다. 본 논문에서는 이러한 문제점을 최소화하여 무선 통신상에서 사용자의 요구에 알맞은 멀티미디어 통계 데이터를 서비스하기 위한 멀티미디어 통계 사서함 시스템을 구축하고자 한다. 제안된 시스템은 통계 데이터들 간의 의미적 관련성에 따라 객체구조와 역파일 구조를 이용한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문을 구성하기 위한 이론적 배경을 제안하며, 3장에서는 무선 인터넷 서비스를 위한 문서 생성 테이블 구조에 대해서 알아보고, 끝으로 결론에 대해서 알아본다.

2. 관련연구

무선인터넷상에서 정보서비스 기술에 대한 연구는 일반 도메인보다는 특정 도메인을 대상으로 연구가 진행되고 있으며, 무선통신이 향후 통신 시장을 주도할 것이라고 판단할 때 무선인터넷 서비스를 위한 특정 도메인에 대한 연구는 매우

중요할 것으로 기대된다. 무선통신 서비스 기술은 데이터 전송, 플랫폼 기반의 서비스 그리고 서비스 접근 방법으로 구분되어진다.

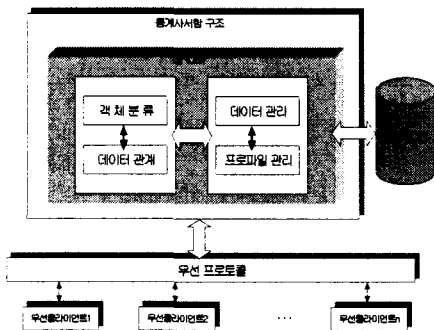
데이터 서비스를 위한 전송기술은 TCP, T/TCP, WTP를 중심으로 하는 전송계층 프로토콜과 무선망에서의 회선교환방식과 패킷교환방식이 주를 이루고 있다[7, 8]. 이들 방법은 무선인터넷 서비스에서 발생하는 낮은 대역폭, 데이터 전송 지연과 불안정한 접속을 문제점과 브라우징을 위한 트랜잭션 위주의 문제점을 해결하기 위하여 다양한 연구가 진행 중에 있다. 특히 WTP는 클라이언트 환경에서도 쉽게 구현이 가능하도록 많은 기능을 제공하고 있으며, 기존의 TCP, T/TCP 프로토콜과 달리 차별화 된 트랜잭션 서비스, 사용자 확인, 트랜잭션 취소와 같은 서비스 기능을 제공하고 있다. 무선통신에서 플랫폼 기반의 서비스는 무선통신 단말기에서 응용 S/W를 다운 받아 사용되는 플랫폼 기반의 서비스 온라인/오프라인으로 연결하여 사용되는 멀티미디어 포털 플랫폼 서비스 등으로 구분되며, 플랫폼 기반의 서비스는 BREW, JAVA, Win CE, MAP 등이 주를 이루고 있으며, 멀티미디어 포털 플랫폼 서비스는 VOD(Video On Demand), AOD(Audio On Demand)을 중심으로 서비스가 이루어지고 있다[2, 3, 5]. 그리고 서비스 접근 방법으로는 서비스를 수행하는 구축언어가 있으며, 웹 기반 구축 언어로는 HDML(Handheld Device Markup Language), WML(Wireless Markup) 등의 언어와 WAP, ME, imode 등의 프로토콜이 사용되고 있다. 그러나 HDML과 WML은 Mobile Device에서 인터넷 접속용으로 개발되어진 언어로서 음성정보 등의 웹방송 서비스에는 이용될 수 없으며, WAP, ME, i-mode 등의 프로토콜은 아직 표준화가 되어 있지 않다는 문제점이 있다. 그밖에 서비스 접근 방법을 위한 연구로는 eNetwork Web Express, Web&, Cyberguide 등의 연구가 있으며, eNetwork Web Express는 무선 인터넷상에서 전송되는 데이터의 크기와 지연을 줄이기 위해서 클라이언트와 서버사이에 위치한 가로채기 기술을 사용하고 있다[9]. Web&은 이동 단말에 대한 정보 서비스 구조로 트랜잭션을 비 동기적으로 수

행하여 웹서비스를 실시하고 있는 모델로서 클라이언트로부터 받은 질의를 Web& 서버가 받아서 이를 처리하고 나서 클라이언트가 원할 때 결과를 제공한다. Cyberguide는 무선네트워크와 이동단말기를 이용하여 각 노드가 위치한 곳의 Locator에 의해서 현재의 위치를 파악하고, 이동단말의 방향성을 탐지한 후 파악된 정보들을 관리 시스템에 의해 서비스하는 시스템 구조이다.

그러나 이들 방법들은 단지 서비스 접근 방법에 대한 구조를 제안했을 뿐 서비스 접근을 위한 객체 관리 기법 등은 제안되지 않고 있다는 점이 문제점으로 지적되고 있다.

3. 통계사서함 시스템 구조

통계사서함 시스템은 도메인 상에서 수집된 데이터를 선별 구성하여 사용자가 요구한 정보를 무선통신상에서 서비스하기 위한 시스템이다. 이와 같은 기능을 수행하기 위한 시스템 구조는 [그림1]과 같다.



[그림1] 전체 시스템 구성도
[Fig.1] The system structure diagram

본 논문에서 제안된 통계 사서함 시스템은 이용자들이 의도한대로 차례를 따라서 서비스를 요구하는 형태가 아니라 시스템과 이용자들이 도메인에서 제공된 데이터를 서비스 받을 수 있도록 구성하며, 사서함 관리를 위한 데이터 관리 및 프로파일 관리, 그리고 이들과 상호작용을 위한 객체분류 및 데이터 관계성으로 구성되며 무선웹서비스를 위한 무선클라이언트 구조로

구성된다. 또한 클라이언트 구조는 사용자 인터페이스, 서비스 처리, 서비스 관리, 클라이언트 메시지 처리 구조로 구성된다.

3.1 객체분류

객체분류는 의미적으로 관련이 있는 통계 데이터들을 클래스로 구성하기 위해 분류하는 작업을 말하며, 본 논문에서는 객체분류를 위해 예측 생성기법과 예측생성을 위해 유사 가중치 방법[6]을 이용하여 객체들간의 유사관계를 결정하게 된다. 예측생성기법과 유사 가중치 부여 방법은 데이터 클래스를 구성하기 위한 방법이며 다음과 같이 정의된다.

<정의1> 예측생성방법

$$P_{c,i} = \bar{r}_a + \sum_{c=1}^n (r_{c,i} - \bar{r}_u) * w_{a,c} / \sum_{c=1}^n w_{a,c}$$

여기서 $w_{a,c}$ 는 유사도 가중치이며, 유사도 가중치는 다음과 같이 정의된다.

<정의2> 유사도 가중치

$$w_{a,c} = \sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{c,i} - \bar{r}_c) /$$

$$\mu_a * \mu_c$$

여기서 c_i 는 클래스 c 에서의 임의의 데이터의 의미하며, a_i 는 임의의 데이터에 대한 attribute를 의미한다. 그리고 \bar{r}_a , \bar{r}_c 는 attribute, class에 대한 순위를 의미하며, μ_c 는 다음과 같이 정의된다.

<정의3> $\mu = \{o \mid C(o) \geq \mu\text{-cut}\}$ 이다.

여기서 $C(o)$ 는 클래스에 포함되는 임의의 객체 정보이며, $\lambda\text{-cut}$ 은 의미적으로 관련이 적은 데이터를 필터링 하는 관련성 정도이다. 예를 들어 2개의 도메인 D_1 , D_2 에서 9개의 통계데이터 d_1 , d_2 , d_3 , d_4 , d_5 , d_6 , d_7 , d_8 , d_9 이 <표1>과 같이 구성되어 있다고 가정하자.

<표1> 유사도에 의한 데이터 관계
 <Table 1> Data Relation by Similarity

d \ c	d ₁	d ₂	d ₃	d ₄	d ₅	d ₆	d ₇	d ₈	d ₉
C ₁	0.10	0.51	0.60	0.41	0.78	0.33	0.00	0.81	0.65
C ₂	0.43	0.91	0.54	0.58	0.88	0.56	0.58	0.79	0.57

여기서 정의3의 μ 를 이용하여 C₁과 C₂에 대해서 유사도를 적용하여 클래스를 구성하면 다음과 같다.

- 유사도를 0.5로 했을 때 데이터클래스는 {d₂, d₃, d₅, d₈, d₉}이며,
- 유사도를 0.6로 했을 때 데이터클래스는 { d₅, d₈ }이며,
- 유사도를 0.7로 했을 때 데이터클래스는 { d₅, d₈ }이 되게 된다.

3.2 데이터 관계성

데이터 관계성은 도메인 상에서 수집된 통계 데이터들에 대해서 데이터를 분류하기 위한 과정으로서 SRR관계성과 DRR관계성에 의해 역파일 구조로 관계성이 표현된다.

(1) SRR(Semantic Reference Relationship) 관계성

SRR 관계성은 프리미티브 데이터들 중에서의 의미적으로 참조 관련성이 있는 데이터를 말하며, 두 개 이상의 데이터가 1:1, 1:m, m:1, m:m의 관계를 유지하여 서로 관련성을 가진 데이터들을 말한다. SRR은 빈도수가 0.5~0.7의 값을 가진 데이터로서 묵시적 관련성, 참조관련성, 하이퍼 관련성으로 구성되며 의미적으로 관련성이 있는 데이터들을 클래스로 구성하기 위해 사용된다.

- 묵시적 관련성(Implicit Relation): 묵시적 관련성은 같은 도메인 상에서 묵시적 관련성을 가진 데이터로서 0.5~0.7사이의 빈도수로 구성된다.
- 참조관련성(Reference Relation): 참조관련성은 임의의 프리미티브 데이터 di와 dj에서 di가 dj를 참조하거나 또는 dj가 di를 참조하는 데이

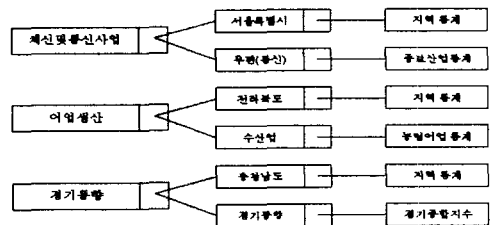
터를 말한다.

- 하이퍼 관련성(HR: Hyper Relation): 하이퍼 관련성은 임의의 프리미티브 데이터 di가 dj에 하이퍼 되는 데이터를 말한다.

(2) DRR(Direct Reference Relationship) 관계성

DRR 관계성은 프리미티브 데이터들 중에서 직접적으로 참조 관련성이 있는 데이터를 말하며, 두 개 이상의 데이터가 1:1, m:m의 관계를 유지하여 서로 직접적 관련성을 가진 데이터들을 말한다. DRR은 빈도수가 0.8~1.0의 값을 가진 데이터로서 명시적 관련성(Explicit Relation), 동치관련성(Equivalence Relation)으로 구성되며 데이터들을 클래스로 구성하기 위해 사용된다.

- 명시적 관련성(Explicit Relation): 명시적 관련성은 같은 도메인 상에서 명시적 관련성을 가진 데이터들로서 0.8~1.0사이의 빈도수로 구성된다.
- 동치관련성(Equivalence Relation): 동치관련성은 프리미티브 데이터와 사상되는 데이터가 동일한 클래스에서 발생하는 데이터를 말한다. 이와 같은 관계성에 따른 역파일 구조는 [그림 2]와 같다.

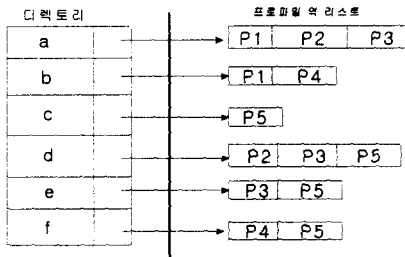


[그림2] 관계성에 따른 역파일 구조
 [Fig.2] Inverted File Structure by Relationship

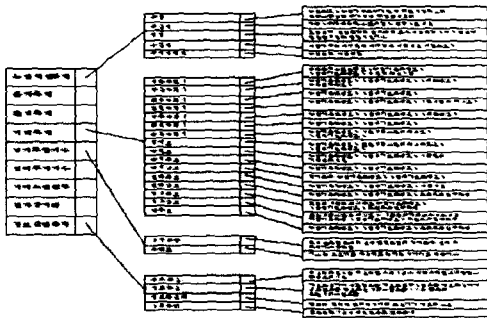
[그림2]와 같은 역파일 구조를 위해 본 논문에서는 카운트 인덱스 구조를 이용하며, 카운트 인덱스 구조는 각 통계 문서 정보에 대하여 관련되는 데이터 정보들을 수집하여 역 리스트로 인덱스를 구성하는 방법이다. 예를 들어 임의의 도메인에 K 개의 관련정보가 있으면 K 개의 프로파일 존재하게 된다. 역파일 구조를 위한 카운트 인덱스 구조는 [그림3]과 같으며, [그림

3]을 이용한 역파일 구조는 [그림4]와 같다.

sample profile		document information (a,c,a,f,b,c)
P1	(a,b)	
P2	(a,d)	
P3	(a,d,e)	
P4	(b,f)	
P5	(c,d,e,f)	



[그림3] 카운트 인덱스 구조
[Fig. 3] Count Index Structure



[그림4] 카운트 인덱스 구조에 의한 역파일 구조
[Fig.4] Inverted File Structure by Count Index

3.2 데이터 관리

데이터 관리는 프로파일 관리와 상호 작용하며, 통계정보 DB와 메모리 상에 존재하는 데이터들을 관리하게 된다. 이때 무선 클라이언트에서 요청된 데이터들의 구조를 파악하여 해당 데이터가 관리 가능한 통계정보인지를 파악하게 된다. 해당 데이터가 관리 가능한 데이터인지는 객체분류, 데이터 관계성 과정을 수행하여 결정되게 되며, 관리가 가능한 문서정보는 클래스를 생성하게 되며, 그렇지 않은 데이터는 비 활성화

화 연산을 수행하여 클래스를 생성하지 않게 된다. 이와 같은 기능을 수행하기 위해서 본 논문에서는 많은 통계 데이터들 중에서 농림어업 통계, 물가 통계, 환경 통계, 지역 통계, 경기종합지수, 설비투자지수, 서비스업활동, 전자상거래, 정보산업통계 등의 기본 데이터만을 이용하여 클래스를 구성하게 된다.

3.4 프로파일 관리

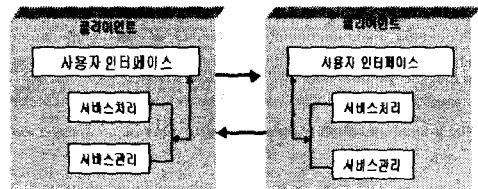
프로파일 관리는 사용자가 데이터에 대한 관심분야를 요구할 때 데이터를 서비스 해 줄 수 있는 관리구조이다. 이때 요구되는 데이터는 프로파일-id와 관리 주기로 구분되며, 프로파일 id는 관심분야의 서비스 종류, 가중치로 구성된다. 그리고 관리주기는 데이터 작성 정보와 데이터 크기 값으로 구성되며 관심분야에 대한 프로파일 정보와 예는 [그림5]와 같다.

프로파일-id	관심분야		관리주기	
	서비스종류	가중치	작성정보	크기
프로파일-id	설비투자지수	0.8	txt 파일	1Mb

[그림5] 관심분야에 따른 프로파일 정보
[Fig.5] Profile Information by Preference area

3.5 시스템 구현

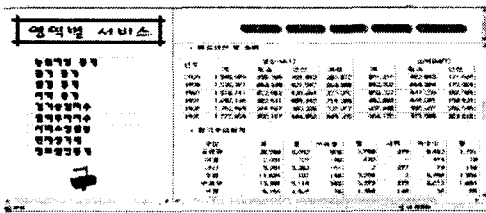
무선 클라이언트는 사서함 시스템과의 서비스 요청을 통해서 사용자 인터페이스와의 상호작용을 통해서 서비스가 수행되며, 이와 같은 서비스 구조는 [그림6]과 같다.



[그림6] 무선클라이언트 서비스 구조
[Fig.6] Service Structure of Wireless Client
사용자 인터페이스는 실제로 웹 기반의 서비스

가 이루어지도록 지원되는 시스템 구조로서 영역별 서비스, 검색을 이용한 서비스, 맞춤 서비스로 구성되며 구현화면은 다음과 같다.

- 영역별 서비스: 영역별 서비스는 이용자가 서비스 받을 영역과 주제 등을 설정하여 서비스를 수행하도록 하며, 이용자의 서비스목표와 내용진도에 따라 서비스 받을 영역을 선택하도록 한다.

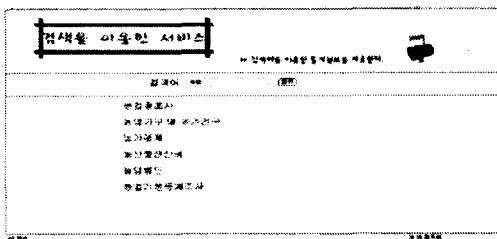


[그림7] 영역별 서비스 화면
[Fig. 7] Service Screen by Area

본 논문에서는 영역별 서비스를 위해 농림어업 통계에 관한 서비스, 물가 통계에 관한 서비스, 환경 통계에 관한 서비스, 지역 통계에 관한 서비스, 경기종합지수에 관한 서비스, 설비투자지수에 관한 서비스, 서비스업 활동에 관한 서비스, 전자 상거래에 관한 서비스, 정보산업 통계에 관한 서비스를 제공한다.

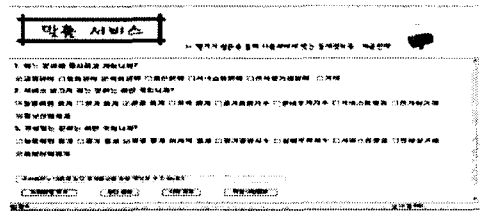
- 검색을 이용한 서비스: 검색을 이용한 서비스는 이용자가 관심 데이터 및 지식이 부족할 때 이용할 수 있도록 도와주는 기능이며, 검색을 이용하여 서비스 받을 주제를 선택하도록 보조해 주는 서비스 기능이다.

이때 사용자가 원하는 데이터를 선택하여 서비스 받고자 한다면 관련 주제어 입력을 통해서 서비스를 수행하도록 보조해 주게 된다.



[그림8] 검색을 이용한 화면
[Fig. 8] Screen using Retrieval

- 맞춤 서비스: 맞춤 서비스 기능은 이용자에게 질문을 통해 이용자가 원하는 서비스를 받을 수 있도록 유도하고 보조하는 서비스 기능이다. 이때 이용자가 원하는 정보만을 보여 주게 된다.



[그림9] 맞춤 서비스 화면
[Fig. 9] Customized Service Screen

4. 결론 및 향후 연구

최근에 정보서비스는 유선인터넷에서 무선인터넷 환경으로 빠르게 변화되고 있다. 정보서비스 관점에서 무선인터넷은 정보전송 등과 같은 여러 가지 제약사항으로 인해 특정 도메인에 대한 서비스가 주를 이루고 있으며, 정보관리를 위한 서비스구조는 미흡한 실정이다. 본 논문에서는 무선통신상에서 사용자가 요구하는 도메인 서비스를 위해 통계 사서함 시스템을 구축하였다. 제안된 시스템은 예측 생성기법과 유사도 가중치를 이용하여 객체 분류기법을 제안하였으며, 데이터 관계성을 이용하여 데이터를 구조화 할 수 있는 역파일 구조를 제안하였다. 또한 분류된 데이터에 대한 관리를 위해 객체 프로파일을 서비스종류, 가중치, 작성정보, 데이터 크기 구조로 관리할 수 있는 구조를 제안하였으며, 이러한 구조를 기반으로 무선 클라이언트에서 영역별 서비스, 검색 서비스, 맞춤서비스가 가능하도록 데이터 서비스 시스템을 구현하였다. 향후 연구를 위해 제안된 시스템이 보다 많은 데이터를 기반으로 다양한 도메인 서비스가 수행될 수 있는 시스템 구축이 필요하다 하겠다.

참고 문헌

- [1] ARC Group, Wireless Internet : Applications, Technology & Player strategies, May, 1999.
- [2] Chabane Djeraba, " Content-Based Multimedia Indexing and Retrieval, " IEEE MULTIMEDIA, 2002, pp. 18-22.
- [3] Guojun Lu, " Techniques and Data Structure for Efficient Multimedia Retrieval Based on Similarity," IEEE TRANSACTION ON MULTIMEDIA, vol 4, No. 3, 2002, pp. 372-384.
- [4] R. Floyd and B. Housel, " Mobile web Access using eNetwork Web Express," IEEE Personal Communications, 1998, pp. 47-52
- [5] Riccardo Leonardi and Pierangelo Migliorati, " Semantic Indexing of Multimedia Documents, " IEEE MULTIMEDIA, 2002, pp. 44-51.
- [6] Prasanna Ganesan, Hector Garcia-Molina, Jennifer Windom, " Exploiting Hierarchical Domain Structure to Compute Similarity, " ACM Transaction on Information Systems, vol. 21, No. 1, 2003, pp. 64-93.
- [7] 김충남, 차세대 무선인터넷 서비스, 전자신문사, 2002.
- [8] 최윤석, 임경식, " 무선인터넷 서비스를 위한 트랜잭션 프로토콜의 구현과 성능평가, " 정보과학회 논문지:정보통신 제29권, 제4호, 2002, pp. 447-458.
- [9] 최인선, 유석대, 오남호, 조기환, " 무선인터넷 망에서 정보 서비스 기술," 정보과학회지,제 18권, 제6호, 2000.6, pp. 39-45.
- [10] 통계 소식지 C, 통계청, 2002, 8
- [11] 통계 소식지 D, 통계청, 2002, 9

이종득

1983년 2월 전북대학교 전산통계학과 졸업(이학사)

1989년 2월 전북대학교 대학원 전산통계학과 졸업(이학석사)

1998년 2월 전북대학교 대학원 전산통계학과 졸업(이학박사)

1992년 3월 - 2002년 2월 서남대학교 컴퓨터정보통신학과 교수

2002년 3월 - 2004년 현재 국립 익산대학 정보통신공학과 교수

관심분야 : 멀티미디어 시스템, 멀티미디어 통신, 무선인터넷, 무선통신 등

김대경

E-Mail : dkkim@mail.chonbuk.ac.kr

1983년 2월 전북대학교 전산통계학과 졸업(이학사)

1985년 2월 동국대학교 통계학과 졸업(이학석사)

1995년 2월 동국대학교 통계학과 졸업(이학박사)

1996년 9월 - 1997년 5월 국립 삼척대학교 교수

1997년 6월 - 2004년 현재 전북대학교 통계정보학부 교수

관심분야 : 멀티미디어 통계, 다변량 해석, 시계열