韓國 컴퓨터情報學會 論文誌
第13卷 第3號, 2008. 5.

2008-13-3-1-8

문서-음성 변환 임베디드 시스템 구축에 관한 연구

이 현 창*, 서 정 만**

A study of Implementing An Embedded System for Conversion from Text to Speech

Hyun Chang Lee*, Jeong Man Seo**

요 약

최신 IT 정보기술 발전에 힘입어 관련 소프트웨어 개발 및 하드웨어 보급이 보편화될수록 IT 기술 활용에 제한적인 사람들은 IT 정보기술 격차를 더욱 크게 느낄 수 있다. 정보통신기기는 일반사용자 이외에도 신체적으로 장애를 가지고 있는 사람들에게 의사소통 및 정보획득을 위한 중요한 수단이 되었다. 또한, 우리나라는 빠르게 고령화 사회로 접어들고 있으면서도 장애를 가지는 이들을 위한 대응 제품을 적기에 제시하지 못하고 있다. 특히, 나이가 들어감에 따라 신체적으로 기능이 저하되는 것 가운데 가장 먼저 시각 기능 저하를 들 수 있다. 시각기능 저하로 인해 장애를 겪고 있는 사람들을 위한 정보전달 수단으로 점자책등이 존재한다. 그러나 일반 서적에 대한 활용과 비교하면 이용 및 활용을 위한 제반 기술이 상당히 부족한 현실이다. 따라서 본 연구에서는 시각기능 저하 혹은 시각장애를 가지는 사람들에게도 일반 서적의 내용을 이해할 수 있도록 리눅스 기반에 소형 스캐너를 부착한 임베디드 시스템 구축에 관한 내용으로서, 문서에 대해 휴대용 스캐너를 활용하여 문자 추출 및 음성으로 변환해주는 통합 시스템에 관하여 살펴본다.

Abstract

According to the development and expansion of software and hardware about recent information technologies(IT), disabled persons in using IT seem to feel more information gap. Devices for IT are important tools for users including disabled persons to communicate with each other and get information. Although the Korea faces ageing society rapidly, products for disabled persons are seldom shown in time for use. As getting older especially, one of the body function disorders is visual disturbance. There are tools, braille lettering, for disabled persons with visual disturbance to communicate or get information from book. Compared to general books, however, braille lettering book is lack of including all of information of our society. Therefore, in this paper, we implement and show an embedded system for disabled person with visual disturbance to get information by scanning text, extracting characters and converting the text to speech automatically.

▸ Keyword : embedded system, text, speech, conversion system

• 제1저자 : 이현창
• 접수일 : 2008. 4. 2, 심사일 : 2008. 5. 6, 심사완료일 : 2008. 5. 24.
* 한세대학교 IT학부 교수 **한국재활복지대학 컴퓨터게임개발과 교수

Ⅰ. 서 론

최신 IT 정보기술 발전에 따라 관련 소프트웨어 개발 및 하드웨어 보급이 보편화 될수록 IT 기술 활용에 제한적인 사람들은 IT 정보기술 격차를 더욱 크게 느낄 수밖에 없다. 이와 같이 일반화된 정보통신기기는 일반 사용자 이외에도 신체적 결함을 가지고 있는 모든 사람들에게 의사소통과 정보획득을 하기 위한 중요한 수단이 되었다.[1]

기술개발이 진행됨에 따라 현대사회의 정보화는 산업현장에서부터 가정에 이르기까지 모든 분야에서 컴퓨터와 인터넷을 통해 다양한 정보를 획득할 수 있으며, 이를 통해 상품과 서비스를 구입할 수 있을 뿐만 아니라 생활의 여유를 즐길 수 있도록 해주는 등 사람들의 생활에 커다란 변화를 가져왔다. 이와 같이 빠르고 다양해지는 서비스의 다변화 속에서 자신의 신체적 결함을 극복하고 보완하기 위한 서비스 활용을 가능케 해주는 IT 정보통신 기기 및 제품들은 이들에게 필수적인 부분이라 할 수 있다.

이와 같은 이유에서 신체적 결함이 존재하는 사람들을 위한 IT 정보기기 개발에 대한 투자 및 기술개발과 보급은 회사 차원에서보다 정부차원의 정책적이면서 경제적 지원이 동반되어 진행되어야 할 것이다. 미국 및 유럽 등 우리나라와 비교될 수 있는 선진국에서는 장애인을 위한 정보통신 활용 및 접근이 장애인들의 삶을 질적으로 향상시킬 수 있다는 인식이 확대되어짐에 따라 정부차원의 장애인용 정보통신 기기 개발과 보급을 위한 지원의 활성화는 우리에게 많은 부분에서 시사점을 주고 있다.[2]

신체적 결함 및 장애 부분들 가운데 시야에 어려움을 갖는 시각장애인들에게 제공되는 정보 전달매체로서 대표적으로 점자책을 들 수 있다. 그러나 점자책도 정보의 홍수인 현대사회의 정보를 모두 수용하지 못하고 있는 실정이다. 이에 본 연구에서는 시각적 결함을 가지는 사람을 위한 문자 인식후 소리로 들려주는 임베디드 시스템을 개발하여 서비스를 제공하고자한다.

먼저, 시각적 결함을 가지는 사람들에게 간편하게 활용할 수 있도록 휴대성을 향상시키면서, 시간과 장소 등에 제약받지 않고, 유비쿼터스 환경[17]에 맞는 언제 어디에서든지 일반인들과 동일하게 책의 정보를 획득할 수 있는 시스템 개발이 필요하게 되었으며, 그 과정으로서 책을 스캐닝을 수행한다. 이후 문자를 추출하고, 추출과정을 통해 문자 인식과정을 수행한다. 다시 인식한 문자를 음성으로 들려주는 시스템으로서 문서를 음성으로 들려주는 시스템 개발이 필요하게 되었다.

문자-음성 변환 시스템을 개발하기에 앞서서 문자 추출을 위한 스캐닝 단계에서 사용자가 휴대하기 편리하게 스캐너의 소형화가 요구되며, 소형 스캐너는 휴대성을 향상시켜 장소에 제약없이 문서를 스캔 후 저장할 수 있게 되었다. 또한, 문자 추출 및 인식 과정을 거쳐서 음성으로 변환하기 위한 음성 합성 기술은 많은 발전을 거듭하여 자연스러운 합성음을 생성할 수 있게 되었다. 이와 같은 제품으로 보이스 텍스트, 매직보잇, 나랏소리, 음성마법사를 들 수 있으며, 국외 제품으로 Emacspeak, Festival, VoiceText 등의 오픈 프로젝트 제품들이 있다.[4,5,6,10,11] 이에 본 연구에서도 문자-음성 변환 임베디드 시스템 개발을 위한 분석 설계 및 구축을 단계별로 살펴보도록 한다.

본 논문의 구성은 다음과 같다. 2장 관련 연구에서는 스캐너와 음성합성에 대해 고찰하고, 3장에서는 문자-음성 변환시스템의 구성 및 기능에 대하여 살펴본다. 제4장에서는 문자-음성변환 시스템 구축 모델에 대한 평가와, 끝으로 5장에서 결론 및 향후 연구 방향에 관하여 살펴본다.

Ⅱ. 관련연구

본 장에서는 기술개발에 필요한 구성 요소에 대해 간략히 살펴본다.

2. 1 스캐너

스캐너는 스캔닝을 수행할 대상에 빛을 반사시킨 다음 빛을 투과시켜 각 부분의 반사도에 대해 데이터를 수집하고 수집된 데이터를 전기 신호로 변환한 다음 디지털화하여 전송 하게 된다.

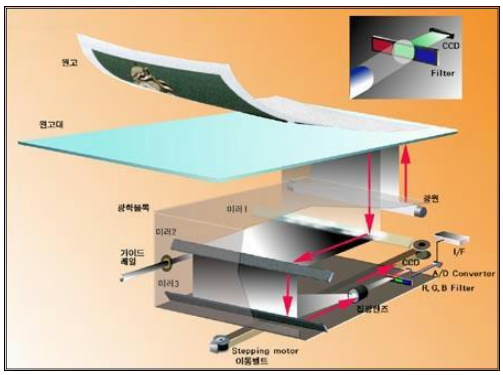


그림 1. 평판 스캐너 동작 과정
Fig 1. Operation processes of plain scanner

〈그림 1〉은 평판 스캐너의 기본 동작 과정을 보여주고 있다. 대부분 원본에서 반사된 빛을 한줄 한줄 측정하고, 반사광을 집광 렌즈를 통해 전하 결합 소자(CCD, charge-coupled device)로 모은다. 해상도는 CCD의 광센서 개수와 스캔하는 라인 간격에 따라 결정된다. 트랜스패런시(투명지) 스캐너는 원본에서 반사되는 빛이 아니라 투과하는 빛을 탐지하여 측정하게 된다.〔1,2〕

2. 2 음성합성

초기의 음성 합성 시스템은 전기적 장치를 이용한 포만트 합성기였으며 두 개의 공진회로를 버저(buzzer)에 의해 여과 되도록 만들어졌다. 이 시스템은 모음의 안정 구간을 두개의 포만트 주파수 즉 F1, F2를 공진 회로로 모델링 하였으며, 이에 따라 모음에 근사화된 합성음이 생성된다. 1939년 벨 연구소의 Dudley는 분석/합성법을 제안해 음성을 인위적으로 가공할 수 있는 방법을 열어 놓았다. 즉 Voder라는 합성 시스템은 유/무성에 따른 음원을 선정하였으며, 피아노 건반과 같은 키보드로 10개의 대역 필터로 구성된 공진회로의 출력을 조절한다. 또한 음의 높이인 기본 주파수를 페달로 조절하게 되어 있다. 이 시스템으로 연속음을 합성하기 위해 오퍼레이터를 1년 이상 훈련시켜야 했다고 한다.

1951년에는 Haskins lab에서 스펙트럼 패턴을 음성으로 변환하는 ^Pattern Playback' 합성기를 개발했다. 이 시스템은 스펙트럼이 그려진 투명한 벨트를 광학적 장치에 의해 읽혀져 음성 발생 과정과 역으로 음성이 합성된다. 1960년 음성 합성은 Fant에 의해 새로운 전기를 마련하게 된다. 지금까지 합성은 주로 시간에 따라 변화해가는 음성의 스펙트럼을 그대로 이용하는 방식이었으나 Fant는 '음성이 어떻게 발생되는가'하는 음향 이론(acoustic theory)에 바탕을 둔 발생학적 모델링을 제안했다. 즉 음성의 발생은 여기 신호와 선형 필터로 모델링(source-filter theory)할 수 있는데 여기 신호는 유/무성으로, 선형 필터는 입술, 구강(oral cavity), 인두강(pharynx cavity)으로 구성된 성도의 공명 효과를 모델링 한다.

이와 같이 source-filter theory에 의한 선형 필터 즉, 성도 전달 함수는 모두 극점(pole)으로 모델링 되며 비음과 같은 음성은 영점(zero)을 첨가해 근사적으로 비강(nasal cavity)을 모델링하게 된다. 이 모델에 따라 최초의 병렬 포만트 음성 합성기인 PAT(Parametric Artificial Talker)와 직렬 포만트 음성 합성기인 OVE I(Orator Verbis Electris)이 개발됐다. PAT 음성 합성기는 세 개의 공진회로가 병렬로 연결되어 있으며 각 공진기의 출력을 단순히 더하여 합성된다. 이 시스템은 세 개의 포만트 주파수, 유/무성 크기, 기본 주파수 등 6개의 제어 파라미터를 갖는다. OVE I은 공진회로가 직렬로 연결되고 두 개의 포만트 주파수, 유/무성 크기, 기본 주파수 등의 제어 파라미터로 모음과 같은 음성만 합성한다. 이후 두 시스템은 파찰음, 비음 등을 모델링한 공진 회로를 추가함으로서 자연스러운 합성음을 생성하게 된다.

1968년 디지털 컴퓨터를 합성에 이용함으로서 전자 회로에 의한 합성기는 쇠퇴하기 시작했다. 그리고 합성 방식도 병렬 포만트 방식과 직렬 포만트 방식이 결합하게 되었고, 유성음화된 마찰음과 같은 세밀한 음성도 모델링 되었으며, 다양한 제어 파라미터도 추가됐다. 1973년에 Holmes는 기존의 병렬 포만트 방식과 음원 모델을 사용해 매우 자연스러운 음성을 생성할 수 있는데, 이 시스템은 1984년에 실시간으로 동작되는 칩으로 구현된 바 있다. 1984년 새로운 음원 모델을 사용한 Infovox SA-101 음성 합성기가 개발됐고, 1985년에는 Klatt가 수학적 모델에 의한 음원을 적용했으며, 1986년에는 Fujisaki가 영점을 첨가한 음원 모델을 제안했다.

한편 성도를 튜브로 단순화해 모델링하고, 이 튜브를 여러 개의 작은 부분으로 세분화한 다음, 공기의 체적 속도나 압력의 분포 등을 전기 회로로 근사화한 조음 합성기

(articulatory synthesizer)가 개발됐다. 이후 이 모델에 유/무성 회로, 비강에 해당하는 회로가 첨가되었고 1975년과 1985년에는 주파수에 따른 영향, 저주파에서 성도 벽면의 움직임이나 성문에서 time-varying impedance를 모델링해 성능을 개선시켰다. 또한 음성 신호 분석/합성에 의한 선형예측 방법이 Itakura, Atal 등에 의해 소개되어 오늘날 대부분의 분석/합성 시스템에 사용되고 있다. 이 방법을 이용해 합성단위를 미리 저장하고 합성시 이 단위를 연결하여 합성하는 방식이 무제한 음성합성기의 주된 방식이다. 최근에는 메모리에 대한 제약이 없어져 시간영역에서 합성하는 TD-PSOLA 방식이 제안되어 요즈음 널리 이용되고 있다.

국내에서는 1990년대 들어 포만트 합성법을 이용한 한국어 규칙합성시스템의 구현에 관한 연구 및 반음절 데이타베이스를 이용한 MPLPC(Multi-Pulse Linear Prediction Coder) 무제한 단어 합성기가 학계에서 개발되었고, 업계에서는 LPC를 이용한 무제한 합성기(명칭: 가라사대)가 PC 플랫폼에서 하드웨어로 개발되어 국내 최초로 상용화된 바 있다. 한편 한국전자통신연구소에서는 LSP(Line Spectral Pairs)와 반음절 데이타베이스를 이용한 합성시스템(글소리-I)을 개발, 이를 교환기의 오디오텍스에 적용한 바 있다. 특히 1992년에 기존의 분석합성법과 다른 TD-PSOLA 방식을 적용하여 매우 명료한 합성음을 생성할 수 있으며, 현재 이 시스템(글소리-II)은 Windows95, UNIX 플랫폼에서 소프트웨어만으로 구동되어 다양한 서비스 개발이 가능하게 되었다.[1]

III. 문서-음성 변환 시스템

본 장에서는 문서-음성 변환 시스템을 구축하기 위해 전체 시스템 구성 및 세부 실제 부품들에 대해 살펴본다.

3.1 문서-음성 변환 시스템 구성

본 절에서는 개발한 문서-음성 변환 시스템의 분석 및 설계 단계로서, 시스템을 구성하는 전체 구성도를 〈그림 2〉에 도시하고 있다.

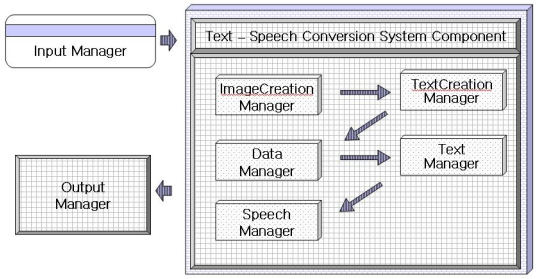


그림 2. 문서-음성 변환 시스템 전체 구성도
Fig 2. System diagram for conversion from text to speech

〈그림 2〉에서 각 기능별 구성을 관리자에 따라 도시하였으며, 세부 내용은 다음과 같다. 먼저, 입력 관리자(Input manager)는 스캐너를 통해서 들어온 데이터 신호를 받아서 이미지생성 관리자(ImageCreation manager)에게 정보신호를 보내준다. 이미지생성 관리자는 전송된 신호를 하나의 이미지 파일로 문서생성 관리자(TextCreation manager)에게 넘겨준다. 문서생성 관리자는 이미지 파일에서 문서를 생성하기 위한 과정을 수행하며, 이러한 과정을 통해 생성된 문서 내에서 문자를 추출해 문서파일을 생성한다. 문서생성 관리자는 순차적으로 생성된 문서파일을 데이터관리자(Data manager)에게 전달하며, 데이터관리자는 생성된 문서 파일에 이어 들어오는 2개의 text파일에서 중복된 부분을 제거하여 한 개의 완성된 문서파일을 생성한다.

데이터 관리자로부터 생성된 완전한 문서파일은 문서관리자(Text manager)로 전달되며 문서관리자는 만들어진 문서를 분석해 문장의 정보와 어절, 억양 등의 정보를 추출해 음성신호 관리자(Speech manager)로 보내준다. 문서-음성 변환 시스템 과정의 마지막 단계인 음성신호관리자는 추출된 정보에 맞는 음성을 합성하여 출력한다. 〈그림 3〉은 상기 〈그림 2〉의 문서생성 관리자, 데이터 관리자, 문서 관리자, 음성신호 관리자의 각 단계별 기능을 모듈형태로 도시한 것이며, 음성으로 변환 출력하기 위한 과정을 기능과 함께 도시하고 있다.

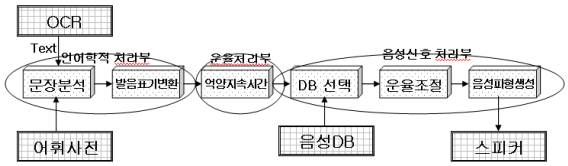


그림 3. 문서에서 음성으로 변환 흐름도
Fig 3. Flow for conversion from text to speech

3.2 문서-음성 변환 시스템 개발 환경 분석설계 및 구축

본 연구에서 개발한 문서-음성 변환 시스템 구축을 위한 구성은 다음 〈그림 4〉과 같으며, 세부 실제 개발 모형을 함께 살펴본다.

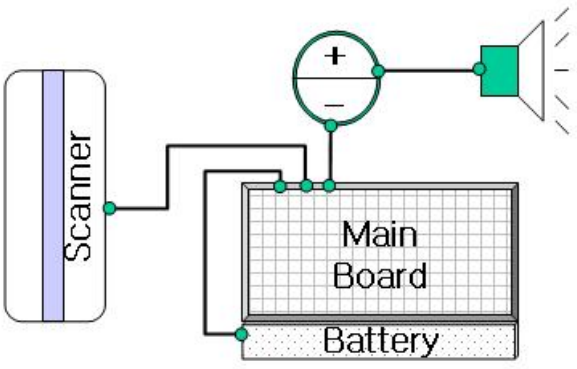
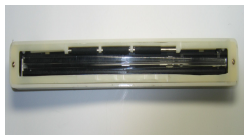


그림 4. 문서-음성 변환 시스템구성도
Fig 4. System diagram of text to speech conversion

〈그림 4〉에서 제시하고 있는 스캐너는 손목에 탈착 가능한 휴대용 스캐너(5.a)를 사용하며, 메인보드(5.b)는 SBC-C677VRL (ACTAC Technology Corp)를 이용하고, 백팩 형식으로 케이스를 제작하여 등에 부착할 수 있게 설계하였다. 메인 메모리는 2기가의 플래시 하드를 사용한다. 배터리(5.c)는 Enine M9600를 이용하고, 이어폰은 귀걸이형 이어폰을 연결해서 사용하며, 시각장애인의 편의를 위해 전원과 동작 2개의 버튼으로 단순화함으로서 편리성을 추구할 수 있도록 구성하고 있다. 볼륨 조절은 기존 사용 방법으로 회전형 버튼을 이용한다. 운영체제는 오픈소스인 리눅스 커널 2.4.1을 이용하며, TTS 시스템은 festival-2.0, OCR 시스템은 gocr-0.40을 사용한다.[1]



가) scanner

나) main board

다) battery

그림 5. 변환 시스템 구성부품
Fig 5. Parts for conversion system

Ⅳ. 시스템 구축 및 평가

최근 IT기술 발전은 정보기기를 이용하여 의사소통과 정보획득을 위한 중요한 수단이 되었다. 이러한 상황에서 신체적 결함 및 장애를 보완하여 정보 활용을 용이하게 해줄 수 있는 정보통신 제품은 장애인들에게 필수적이다. 이에 본 장에서는 개발한 문서-음성 변환 시스템에 대해 분석 및 설계를 통해 얻을 수 있는 장단점을 살펴보고자 한다.[1]

4.1 문서-음성변환 시스템의 평가 분석

본 연구에서 개발한 문서-음성변환 시스템은 신체 일부에 부착된 스캐너를 이용하여 문서를 스캐닝한다. 스캐닝한 결과로 생성되어진 이미지 문서에서 글자를 추출하여 각 글자에 상응하는 음성으로 시각장애인에게 실시간으로 음성 서비스를 제공할 수 있게 된다. 이와 같은 절차를 통하여 구축되는 시스템은 스캐너, 문자추출, 문서의 음성 변환 등의 기능이 하나의 시스템에 통합된 시스템이기 때문에 각각의 기능별 장치를 사용하는 것보다 실시간이면서 빠른 응답을 얻을 수 있게 된다. 또한 개인 휴대성이 뛰어나기 때문에 백팩의 형식으로 부착이 가능하여 일상 생활 및 활동에도 지장 없이 언제 어디에서든지 활용이 가능하다.

이와 같은 장점을 지니고 있지만 스캔하는 속도, 기울기 등의 영향이 스캔율에 영향을 크게 끼치기 때문에 스캔율이 좋지 않을 경우 문자판독 또한 정확성이 떨어지게 된다. 그 외에 해결되어져야 할 사항 중에 TTS의 발음이 자연어를 사용하는 인간의 발음에 크게 미치지 못하기 때문에 이에 대한 기술 개발이 필요하며, 본 시스템의 특징을 표 1과 같이 요약할 수 있다.

표 1. 특성 비교
Table. 1 Comparison of main feature

항목	기존시스템	제안시스템
문서>음성 통합기능 제공도	중	상
휴대성	중	상
실시간제공	하	상
문서추출의 정확성	상	중
문서 판독율	상	중
음성의 정확성	상	중

표 1에서와 같이 기존시스템은 휴대성이 불편하며, 실시간 제공이 떨어지는 반면, 제안 시스템은 휴대성이 편리하며, 통합기능이 우수하고, 통합된 문서 음성 변환 서비스를 실시간으로 제공할 수 있는 장점을 지니고 있다.

4.2 기대효과 및 응용분야

본 연구에서 제시하고 있는 문서-음성 변환 시스템은 시각장애인들이 타인의 도움 없이 일반인과 유사하게 일반 문서를 읽을 수 있게 되어 더욱 많은 정보를 손쉽게 획득 할 수 있게 서비스를 제공해주고 있다. 더욱 특징적인 부분은 일반인들이 오히려 어두운 부분에서 책을 판독하기 어렵지만 본 연구의 결과를 응용할 경우 어두운 공간에서도 문서에 대한 판독을 실행할 수 있다. 그 외에도 인쇄물의 빠른 문서화로 전자책, VoiceBook 등을 제작할 수도 있으며, GPS와 연결하여 음성 네비게이션을 구축할 수도 있게 된다.

V. 결론 및 향후 연구방향

신체적 결함을 가진 사람들 가운데 시각장애인들에게 정보 전달매체로 점자책이 있지만 일반 책에 비해 턱없이 부족한 것이 현실이다. 이에 시각장애인이 휴대할 수 있으면서, 시간과 장소에 제약받지 않고, 언제 어디서든 일반 책을 스캐닝 이후 글자 추출과 추출된 글자를 음성으로 읽을 수 있게 해주는 문서-음성 변환시스템의 개발이 필요하게 되었다.

이에 본 연구에서는 문서-음성 변환 임베디드시스템 구축을 통해 향후 개인 맞춤형 및 신체적 결함을 가지는 사용자를 위한 서비스를 제공할 수 있는 프로토타입을 개발하여 이에대해 살펴보았다. 다음 그림은 실제 개발된 시스템의 시연을 보이고 있는 모습이다.

그림 6. 문서-음성 변환 시스템의 시연
Fig 6. Preview for conversion system

<그림 6>에서 휴대의 편리성을 위해 배낭 부분에 보드를 넣어서 우측 손으로 스캐너를 스캐닝하는 과정을 시연하는 모습이다. 향후 본 시스템에 대해 휴대성 및 디자인적인 측면이 강조된 연구와 문자 인식률 향상과 음성이 사람과 유사한 음성 출력에 관한 연구를 지속할 것이다.

참고문헌

[1] J. H. Kim W.H. Lee H.C. Lee, "Design and Analysis of T2S System Implementation", KIPS Conference on, Vol.13, November 2006.
[2] http://ksc.digitalsme.com/club/club_board_view.
[3] http://blog.naver.com/pr2780?Redirect=Log&log
[4] http://www.ubiu.com
[5] http://www.actac.co.kr
[6] http://www.maxan.com
[7] http://www.actac.com.tw/Show_product.asp?id=623
[8] http://kelp.or.kr/korweblog/stories.php?story=05/10/20/4723113
[9] http://kelp.or.kr/korweblog/stories.php?story=04/04

/28/7166008
〔10〕 http://www.aleph1.co.uk/taxonomy/term/31/
〔11〕 http://www.linux-usb.org
〔12〕 http://www.linux-mtd.infradead.org
〔13〕 http://blog.naver.com/come2alex/10000941689
〔14〕 http://www.epson.co.kr
〔15〕 이정철, 이상호, "교육용 한국어 TTS 플랫폼 개발", 말소리 제 50호, p41-50, 2004
〔16〕 한국과학기술원, "한국어 텍스트에서의 문장구조 추출 도구 개발"
〔17〕 이현창,김정곤. "u-헬스 케어 환경에서 뇌혈관 질환 진단 모델 연구." 한국컴퓨터정보학회논문지, 제11권 제6호, pp107-112, 2006.12

저 자 소 개

이 현 창

2001년 홍익대학교 박사
2001년 경인여자대학 교수
2003년 한세대학교 교수 재직
데이터 웨어하우스, 영상처리, 유비쿼터스 컴퓨팅

서 정 만

2003년 충북대학교 컴퓨터공학과 박사
2002년 ~ 현재 한국재활복지 대학 컴퓨터게임개발과 교수
관심분야 : 실시간처리, 게임프로그래밍, 가상현실, 데이터베이스