

# 자연어처리를 이용한 구조물 설계지식정보 데이터베이스 구축에 관한 연구

A Study on the Construction of Database contains Knowledge  
for the Structural Design using the Natural Language Processing

이 민 호\* · 이 정 재 · 김 한 중 · 윤 성 수(서울대)

Lee, Min Ho\* · Lee, Jeong Jae · Kim, Han Joong · Yoon, Seong Soo

## ABSTRACT

In this study, by using the natural language processing of the field of artificial intelligence, automated index was performed. And then, the Natural Language Processor for Constructing Database (NALPDB) has been developed. Furthermore, the Design knowledge Information RELational DataBase (DIREDB) has been also developed, which is designed to interlock the knowledge base. DIREDB processes both the documented design-data, like a concrete standard specification, and the design knowledge from an expert. DIREDB is also simulates the design space of structures accordance with the production rule, and thus it is determined that DIREDB can be used as a engine to retrieve new knowledge and to implement knowledge base that is necessary to the development of automatic design system.

## 1. 서론

구조물의 설계에는 안전성, 사용성, 경제성 등의 조건을 만족하기 위한 기준이 필요하다. 발확실한 물리적·재료적 특성을 가진 철근콘크리트 구조물 설계에서는 설계기준을 만족시키기 위해 지금까지 대부분 전문가의 경험적 지식이 중요한 역할을 수행하였다.<sup>2,11,12)</sup> 그러나, 전문가의 경험적 지식은 감각적이고 비정형적인 성격이 강하고 문장으로 표현되어 있지 않기 때문에 전산화하기 어려운 반면, 콘크리트 표준시방서와 보고서 같은 문헌자료는 설계지식을 객관적이고 정형적으로 표현하고 있고 문장으로 표현되어 있어 상대적으로 전산화가 용이하기 때문에, 문헌지식을 전산설계시스템에 효과적으로 이용하려는 연구가 필요하다.<sup>7,13)</sup>

전문가에 의한 설계는 경험, 지식, 직관을 바탕으로 설계 대상 구조물의 전체적인 특징을 종합적으로 고려하여 설계를 하지만, 단순 반복적인 수치계산과 대안의 생성과정에서 다양한 오류가 발생할 수 있으며 시간적 제약을 많이 받는다. ADOSS<sup>11,20)</sup>와 SAFE<sup>17)</sup> 등의 전산설계시스템은 이런 문제점을 해결하여 단순 반복적인 수치계산을 정확히 수행하고, 짧은 시간에 다양한 설계 대안을 사용자에게 제시하였다.<sup>12,17)</sup> 그러나, 과거의 전산설계시스템은 프로그램된 작업순서에 의해서만 설계 과정을 검토하기 때문에 구조물의 특징을 고려한 종합적인 판단을 할 수

없었다. 전문적인 지식과 경험을 바탕으로 결론을 도출하는 전문가처럼 규칙으로 표현된 전문적인 설계지식정보를 이용하여 설계를 수행할 수 있는 전산설계 전문가시스템이 개발되었다.

전문가시스템은 많은 전문가로부터 모니터링(monitoring)하여 지식을 획득하였기 때문에, 시스템을 구성하는 지식의 축적에 많은 시간이 소요되며, 획득된 지식을 전산시스템에 이용되는 규칙의 형태로 표현하기 어려웠다. 그러나, 구조물 설계의 증가로 인해 전문가의 수가 절대적으로 부족했기 때문에, 초보 기술자에 의한 설계가 가능한 HI-RISE<sup>11,18)</sup>와 SACON<sup>11,19)</sup> 등의 전문가시스템의 개발이 활발히 이루어졌다.<sup>11,14,15)</sup> 전문가시스템은 설계시스템 구성이 복잡하고 일단 시스템이 구성되고 난 후 수정이 어려우며, 급격히 증가하는 지식을 관리할 수 있는 지식기반시스템의 기능이 미약하여 뚜렷한 발전을 거두지 못하였다.<sup>5,6,11,13)</sup> 이러한 새로운 응용분야에 있어서의 지식베이스 특성은 대량의 공유가 가능한 분산된 지식과 다양한 형태의 지식을 포함하고 있다. 이런 문제를 해결하기 위하여 다양한 대량의 자료관리에 유리한 LKBS<sup>6)</sup>와 EDUCE<sup>6,18,19)</sup> 등의 데이터베이스 관리시스템과 지식기반시스템을 결합한 시스템이 개발되었다. 그러나 이런 결합시스템은 하나의 시스템에 다른 시스템의 기능을 포함하기 때문에, 시스템의 효율을 저하시키는 결합 불일치(impedance mismatch)가 발생할 수 있다.

본 연구는 설계에 이용되는 콘크리트 표준시방서와 보고서 등과 같은 문헌상의 객관적인 지식뿐만 아니라, 문장형태로 표현 가능한 전문가의 경험적 지식을 자연어처리를 통하여 분석하였다. 분석된 자료는 설계지원 전문가시스템 등에서 규칙으로 사용될 수 있도록 개발된 설계지식정보 데이터베이스(Design knowledge Information RELational DataBase, DIREDB)에 구축하고, 철근 콘크리트 구조물을 대상으로 DIREDB의 적용성을 검증하고자 한다.

## 2. 설계지식정보 데이터베이스 설계

### 2.1 설계지식정보의 구성

#### 1) 서술형 설계지식정보

문장으로 표현된 설계지식정보는 명제(proposition)와 같이 사실로 표현할 수 있는 논리형 정보와 일련의 행동에 따른 결과로 표현할 수 있는 규칙형 정보로 구성되어 있다. 논리형 정보를 이용하여 사실을 표현할 때는 논리 베이스와 관계 모델(logic-base and relational model)을 사용하며, 절차형 규칙 표현은 문제를 해결하는 방법의 지식을 저장하기 위해 규칙 베이스 모델(rule-base model)을 사용한다.<sup>6,18)</sup> 논리형 정보에 의한 설계정보 표현은 사실(facts)과 단언(assertions)이며, 수학이나 논리학에서 사용되었던 논리를 이용하는 명제로서 수식으로 표현할 수 있다. 규칙형 정보는 가정(if)과 결론(then)으로 표현되는 설계지식정보이며, 결정을 내리는 상황이 요구되는 경우에 제약조건, 경계값, 범위가 주어진 경우 등에 유용하게 사용되어진다.

#### 2) 수식과 메타지식정보

수식과 메타지식정보는 설계에 이용되는 구조물의 기하학적 구성요소, 단면 및 재료, 하중조건 등을 표현하거나 설계가 이루어진 배경, 주요제약 조건 등의 통계 및 실험자료들의 결과를 정리한 것들로 구성된다. 수식과 메타지식정보는 일정한 형식을 지니고 있기 때문에 설계지식정보 데이터베이스에 자료 구축이 비교적 쉽다. 구조물을 설계할 때 이용되는 수식과 메타지식정보는 구조물 구성요소로 보, 슬래브, 기둥, 벽체 등 그 대상이 한정되어 있고 각 요소들도 단면정보, 하중정보, 부재력정보 등을 공통으로 가지고 있으며 설계에 없어선 안될 자료들이므로 일반설계 프로그램 작성시 주 이용 대상이 된다.

## 2.2 기호 처리

개별 어휘를 중심으로 문장을 분석하는 것은 자연어처리 분야의 여러 응용에서 매우 바람직한 방법으로 인식되었다. 정보 검색을 위하여 주어진 문장 중 중요어휘(keyword)나 중요분맥을 추출하는 문제인 경우, 문장 전체에 대해 성공적인 분석 결과를 얻을 수 없더라도, 그 문장을 이루는 독립적인 어휘들을 인식할 수 있으면 부분적인 정보를 얻을 수 있다.<sup>3,8,9,10,16)</sup> NALPDB는 입력 문장으로부터 기호를 분석하고, 분석된 기호를 정의된 사전들을 통해 변환할 수 있기 때문에, 이들을 조합하여 의미 있는 내용을 생성할 수 있도록 개발되었다.

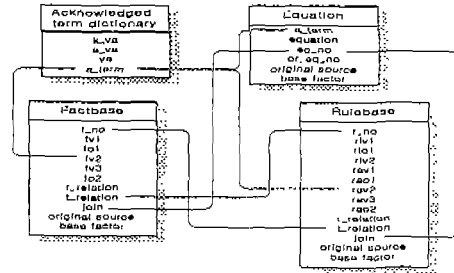
## 2.3 설계지식정보의 자료구조 정의

구조를 설계와 관련된 설계지식정보는 서술형 설계지식정보, 수식, 메타지식정보, 기호 등으로 분류하며, 각 요소들간의 원활한 운용과 참조 및 자료의 일관성 유지를 위해 서술형 설계지식정보와 메타지식정보는 지식베이스 테이블, 수식은 수식테이블, 기호는 사전테이블로 각각 정의하였다. 지식베이스 테이블은 논리형 정보와 메타지식정보를 저장하는 사실베이스 테이블과 규칙형 정보를 저장하는 규칙베이스 테이블로 구성되며, 사전 테이블은 대표어사전(acknowledged term dictionary) 테이블 테이블로 구성된다. <Table1>은 각 테이블의 주요 기능을 나타내고 있다. <Fig.1>은 테이블의 스키마(schema)간의 관계를 전체적으로 나타내었다.

Table.1 Main functions of each tables

Table	Main functions
knowledge base	save facts and rule information, joined equation table
equation	joined fact table and rule table, refer to number given to an equation
dictionary	organized acknowledged term, used pre_ and post_processor

Fig.1 Relation chart of the defined tables



## 2.4 설계지식기반의 일관성 유지

주어진 상황이 동일할 경우, 항상 같은 결론을 추론해 낸다면 지식에 일관성이 있으며, 그렇지 못하면 비일관성이 존재한다. 지식의 비일관성은 구축할 지식의 양이 많아지고, 다양한 문헌자료로부터 설계지식을 구축할 경우 서로 다른 이론을 수용한 문헌자료나 전문가의 경험적 지식에 의한 오류에 의해 발생하고, 새로운 연구결과에 의해 지식의 내용이 변하여 그 변경된 내용을 새롭게 설계지식기반으로 구축할 경우에 발생할 수 있다.

비일관성의 종류에는 중복규칙(redundant rule), 모순규칙(contradiction rule), 충돌규칙(conflict rule), 포함규칙(subsumption rule), 불필요한 전제부(unnecessary condition) 등이 있으며, 지식의 비일관성은 구축된 설계지식정보 데이터베이스를 전문가시스템에 이용할 경우에 예측치 못한 결과를 초래할 수 있기 때문에 제거되어야 한다.

## 3. 자연어처리에 의한 설계지식정보 데이터베이스 구축

3장에서 2장에서 제시한 설계지식정보 데이터베이스 구축을 위한 자료구조 및 테이블의 정의를 이용하여 자연어 처리시스템인 NALPDB(NATURAL Language Processor for constructing DataBase)를 개발하였다. NALPDB의 순서도는 <Fig.2>와 같다.

### 3.1 대표어 변환

#### (1) 대표어사전 작성

자연 언어로 표현된 설계정보를 처리하기 위해서는 하나의 기호를 여러 가지로 표현하고 있는 표기법을 통일해야 한다. 컴퓨터는 프로그래밍 그대로 작업을 처리하기 때문에, 동일한 기호에 대하여 표기법이 상이하다면 결과도 역시 상이할 것이다. 이것은 자연 언어를 이해할 수 있는 컴퓨터가 자연 언어에 존재하는 다양한 표현 및 실제 의미와는 다르게 표현되는 패턴(patterns)에 대한 인간의 프로그래밍 능력에 의존하기 때문이다. 대표어사전은 자연어처리 시스템의 효율성을 높이며 신뢰성 있는 분석 결과를 얻기 위하여 농업토목구조물 설계에 주로 이용되는 콘크리트 표준시방서의 기호와 설계변수를 중심으로 작성하였다.

#### (2) 한글 자연어처리의 전처리

입력된 설계지식정보 문장은 일상적으로 사용되는 각종 부호들과, 영어, 한자, 특수기호 등의 다양한 표기들이 포함되어 있기 때문에, 먼저 이들을 구분해야 한다. 또한, 데이터베이스에 입력된 자료가 재사용되기 위해서는 전처리를 통해 표기법을 통일하는 것이 중요하다. 본 연구에서는 용어와 기호 등의 표기법을 통일하기 위해 자연어 전처리를 개발하였다. 개발된 자연어 전처리의 알고리즘은 <Fig.3>과 같다.

### 3.2 HAM을 이용한 자동 색인 및 서술형 설계지식정보 전처리기 개발

본 연구에서는 HAM(Hangul Analysis Module)<sup>1)</sup> 중에서 형태소 분석을 기반으로 하는 자동 색인(automatic indexing) 기법을 사용하였다. 자동 색인시 발생할 수 있는 오류를 해결하기 위하여 다음과 같이 서술형 설계지식정보의 pre\_processor를 개발하였다.

#### (1) 복합명사화

한국어와 같이 조사가 발달된 언어의 경우에는 한 어절에서 조사를 제외한 나머지를 색인어로 채택한다. 이 방법이 가지고 있는 가장 큰 문제는 복합 명사에 대한 처리이다. 예를 들어, ‘철근항복강도’이라는 복합 명사는 “철근항복강도에 대한...”에서와 같이 붙여 쓸 수도 있고, “철근 항복강도에 대한...”과 같이 띄어 쓸 수도 있다. 전자의 경우에는 ‘철근항복강도’가 색인어로 채택되어 분석되지만, 후자의 경우에는 ‘철근’과 ‘항복강도’가 각각 색인어로 채택되어 분석되므로 색인을 한 이후에도 이에 대한 처리가 필요하다. 사용자는 복합명사화창을 통해 선택한 둘 이상의 색인어를 하나의 복합명사로 만들 수 있다.

#### 2) 미등록어 변환과 삭제

미등록어는 자동 색인을 통한 설계지식정보 문장의 분석에서 분석 결과로 출력되지 않을 수도 있기 때문에, 미등록어를 필요에 의해 대표어사전에 등록될 단어로 변환해야 한다. 사용자는 사전에 등록된 단어를 변환창에 입력된 단어를 기본으로 검색할 수 있고, 변환창에 직접 입력을 통해 검색할 수 있다. 또한, HAM을 이용하여 자동 색인을 하면, NALPDB는 일차 색인어와 후보 색인어를 모두 출력한다. 사용자는 삭제창을 의미가 없고 중복된 색인된 색인어를 필요에 따라 삭제하기 위해 사용한다.

#### 3) 입력 자료형 결정

문장으로 표현된 설계지식정보는 논리형 정보와 규칙형 정보로 나눌 수 있다. 사용자는 인터페이스를 통해 자료형을 결정하며, 자료형 결정창은 단순히 입력된 문장이 사실베이스에 구축될 논리형 정보인지, 규칙베이스에 구축될 규칙형 정보인지에 대한 결정을 하는 기능을 한다.

Fig.2 Flow chart of NALPDB

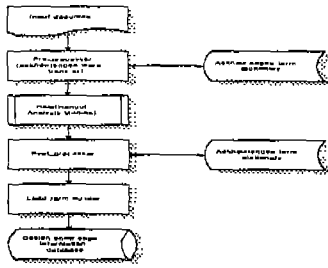
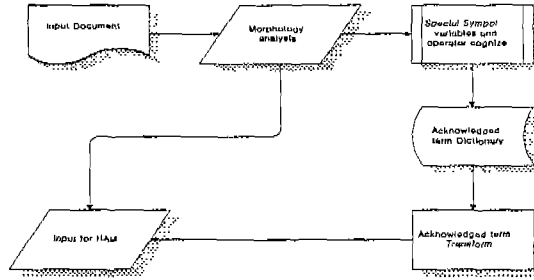


Fig.3 Flow chart of natural language pre\_processor



4) 자동입력기 개발

서술형 설계지식정보 중에서 하나의 문장이 두 개 이상의 지식을 나타낼 수 있다. 외부자료의 입력기능에 의해 동시에 많은 서술형 설계지식정보가 입력되어 DIREDB에 구축될 경우, 입력된 자료를 순차적으로 처리하여야 한다. 자동입력기는 입력된 문장으로부터 색인된 다양한 형태의 설계지식정보 자료를 DIREDB에 구축할 수 있다.

3.3 적용 및 검증

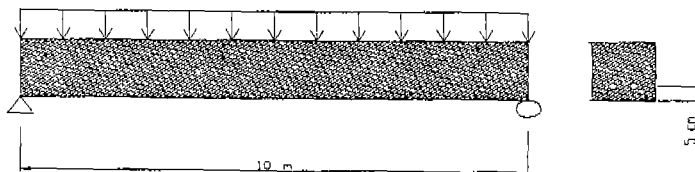
본 연구에서는 객체관계형 데이터베이스 관리시스템인 PostgreSQL 6.5을 이용하여 DIREDB를 구축하였으며, <Fig.4>의 구조물을 대상으로 적용성 및 검증을 실시하였다.

전문가는 구조물 설계를 위해 자신의 경험 지식과 콘크리트 표준시방서와 같은 문헌자료를 이용할 것이며, <Fig.4>의 구조물에 대하여 <Table2>와 같은 순서로 설계하였다. <Table2>는 10명의 전문가들의 설계과정을 정리한 것이다.

DIREDB에서 <Table3>과 같은 순서에 의해, <Fig.5>과 같은 설계공간(design space)을 구성하여 구조물을 설계하였다. 입력된 초기조건에 의해 DIREDB로부터 자료를 추출하였고, 그 결과는 <Table4>에 정리하였다.

<Table2>에 사용된 서술형 설계지식정보, 수식, 메타형 설계지식정보는 <Table4>에서 모두 재현되었기 때문에, 본 연구에 의해 구축된 DIREDB는 철근콘크리트 구조물 설계지식정보 표현과 설계지식정보의 저장·관리 지원이 가능할 것으로 판단된다.

Fig.4 Design example



Design the cross-section, when  $\sigma_v = 3500 \text{ kgf/cm}^2$ ,  $\sigma_{ck} = 300 \text{ kgf/cm}^2$ ,  $W_L = 1.4 \text{ tf/m}$ .

Table2 Design procedure of expert

Order	Detail design	Reference
cross-section assumption	assumption of 'd' assumption of 'b'	$d/b = 15 \sim 25$
load computation	computation of dead load combination of load	table 3.1 equation 3.9
steel ratio assumption	computation of $D_{max}$ & $D_{min}$	table 4.3.1 equation 4.3
normal strength computation	computation of ' $A_s$ ' computation of 'a'	$\rho = A_s/bd$ $a = A_s \cdot \sigma_y / (0.85 \cdot \sigma_{ck} \cdot b)$
trial method	strength reduction coefficient( $\phi$ ) strength compare	$M_u \leq \phi \cdot M_n$

Fig.5 Design space

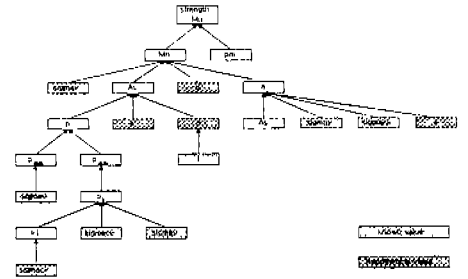


Table3 Working flow of manually query(part)

Order	Constraint	Query	Searching value	Next query
1	Boundary condition Input condition(strength)	Mu	compare & trial method	phi, Mn
1.1	consider base factor	phi	constant	
1.2	hypothetical value	Mn	d	As, d a
1.2.1	hypothetical value	As	b, d	rho
1.2.1.1	hypothetical value	rho	rho	Dmin, Dmax
1.2.1.1.1	consider base factor	Dmin	operation	constant
1.2.1.1.2	consider base factor	Dmax	operation	rho
1.2.1.1.2.1	consider base factor	rho	operation	k1
1.2.1.1.2.1.1	input condition	k1	operation	
1.2.2	hypothetical value	d	operation	l
1.2.2.1	input condition	l	constant	
1.2.3	operation	a	operation	

Table4 Searching data from the DIREDB

Search	Knowledge base output	Equation output
Mu	f29	e12
Mn	f9	e1
phi	f11	
As		e9
a		e3
d	f25, f27, f30	
b		
rho		e8
Dmin	f15	e7
Dmax	f27	
rho		e1
k1	f13, r6, r7	e10

- f-: knowledge of the fact base table
- r-: knowledge of the rule-base table
- e-: knowledge of the equation table

#### 4. 결론

본 연구에서는 문장으로 표현할 수 있는 설계지식정보를 인공지능(artificial intelligence) 분야의 하나인 자연어처리를 이용하여 자동색인을 하고 색인된 결과의 처리를 위해 NALPDB를 개발하였다. 처리된 자료는 지식베이스와 연동이 가능하도록 정의된 DIREDB에 구축하였으며, 그 결과를 요약하면 다음과 같다.

1. 자연어처리를 이용한 자동색인을 통하여 콘크리트 표준시방서와 문장으로 표현된 전문가의 경험 지식으로부터 설계지식정보를 추출하였다.
2. 추출된 설계지식정보는 NALPDB를 통하여 전문가시스템에 이용되는 지식베이스와 연동이 가능하며 자료의 유지·보수가 쉬운 DIREDB에 구축하였다.
3. DIREDB로부터 추출한 자료는 전문가에 의한 설계결과와 비교하여 설계지식정보의 높은 재현성이 있었다.

앞으로 DIREDB로 구축할 설계지식정보의 범위를 확대하고, 영어와 같은 타 언어에 대한 대표어사전을 구축한다면, 본 연구에서 개발한 시스템의 적용 영역이 확대될 것으로 기대된다. 그리고, 사용자 인터페이스 부분에 대한 완전 자동화가 이루어지고, DIREDB로부터 추출된 자료를 자동으로 규칙으로 변환할 수 있는 시스템이 개발된다면 통합설계자동화에 기여할 수 있을 것으로 판단된다.

## 5. 참고 문헌

1. 강승식, 1996, HAM : 한국어 형태소 분석 라이브러리, 한글공학연구소
2. 고일두, 김문현, 최창근, 1983, 컴퓨터를 이용한 RC 부재 자동설계에 관한 연구, 건설기술고도화 시스템의 개발에 관한 연구(I), 과학기술원
3. 김영택, 1994, 자연 언어 처리, 교학사
4. 대한토목학회, 1996, 콘크리트 표준시방서, 건설교통부
5. 류은상, 1995, 데이터베이스를 이용한 전문가시스템의 효율적인 지식베이스 구축에 관한 연구, 국방대학원 석사학위 논문
6. 민미경, 1993, 대규모 지식베이스의 운용 : 지식 표현, 매칭, 저장 관리 기법, 서울대학교 박사학위 논문
7. 박익수, 정성구, 신태양, 신남수, 유영동, 1998, 하이퍼미디어 지원 지능형 건축설계 학습 시스템에 관한 기초적 연구, 대한건축학회논문집 14권 1호, p61-70
8. 서울대학교 시스템공학연구소, 1996, 영한 한영 텍스트 자동 번역 기술 개발, 정보 통신부 연구개발 사업 최종 연구보고서
9. 우요섭, 1992, 데이터베이스의 자연언어 인터페이스, 한양대학교 박사학위 논문
10. 임권복, 1995, 한국어 형태소 분석에서의 오분석 제거와 중의성 해결, 연세대학교 박사학위 논문
11. 장주홍, 윤성수, 김한중, 이정재, 1999, 구조물 설계에 있어서 데이터베이스 관리 시스템의 응용에 관한 연구, 한국농공학회지 3권 8호
12. 정종현, 이재철, 김치경, 홍성목, 1995, 바닥판 시스템 설계자동화를 위한 데이터베이스, 대한건축학회 춘추학술발표대회 논문집, p45-50
13. 조문상, 1996, 건축 설계 지식베이스 개발을 위한 Knowledge Acquisition 방법에 관한 연구, 대한건축학회논문집 12권 6호
14. 홍성목, 허명제, 고일두, 김치경, 1989, 인공지능을 이용한 구조설계 전문가시스템에 관한 연구(I), 대한건축학회논문집 5권 3호
15. 홍성목, 허명제, 고일두, 김치경, 1989, 인공지능을 이용한 구조설계 전문가시스템에 관한 연구(II), 대한건축학회논문집 5권 4호
16. Brill E., R. J. Mooney, 1997, An overview of empirical natural language processing, AI Magazine vol18 no4 p13-24
17. Habibullah A., 1988, SAFE(Slab Analysis by the Finite Element Method), Computers & Structures Inc
18. King D., 1990, Intelligent Decision Support: Strategies for Integrating Decision Support, Database Management, and Expert System Technologies, Expert Systems with Applications vol1 no1 p23-38
19. Sellis T., 1990, A homogeneous Approach to Bringing AI and DBMS Together, Proceedings of the CompSac p32-33
20. Sheth A., A. A'hare, 1991, The Architecture of Braid: A System for Bridging AI/DB Systems, Proceedings of the 6th International Conference on Data Engineering p570-581