

## FCA를 이용한 임상서식지의 분류와 체계화

김명기, 황석형<sup>1</sup>, 김홍기, 강유경<sup>1</sup>, 최희철, 김동순<sup>1</sup>

서울대학교 치과대학, 선문대학교 컴퓨터정보학부 온톨로지기반 소프트웨어공학연구소<sup>1</sup>

### Classification and Conceptualization of Clinical Documents using Formal Concept Analysis

Myeng-Ki Kim, Suk-Hyung Hwang<sup>1</sup>, Hong-Gee Kim, Yu-Kyung Kang<sup>1</sup>, Hee-Chul Choi, Dong-Soon Kim<sup>1</sup>

DERI Seoul, School of Dentistry, Seoul National Univ.; Div. of Computer and Information Science, Sunmoon Univ.<sup>1</sup>

#### Abstract

**Objective:** Ontology is becoming a core research field in the realm of medical informatics. The objective of our ongoing research is to explore the potential role of Formal Concept Analysis(FCA) in a context-based ontology building support in a medical domain. The concept hierarchy plays an important role as the backbone of ontology, but its construction is a complex and time-consuming process. We present a novel approach to the automatic acquisition of taxonomies or concept hierarchies from clinical documents. **Methods:** Our approach is based on FCA, a mathematical tool used in data analysis and knowledge engineering. It provides methods to group objects and attributes into concepts, pairs of object-sets(clinical documents) and attribute-sets(fields contained in the clinical documents), such that the binary relation can be presented in a concept lattice. Based on the FCA, we have applied our approach for 8 clinical documents used in a university hospital. As a result of our experiments, we can extract 15 concepts with 7 common fields that can be shared with 8 clinical documents. **Results:** We show how FCA can be used to classify clinical documents and acquire a concept hierarchy for the medical domain out of the clinical documents with maximal property factorization. **Conclusion:** The whole of our work is based on the concept lattice of which allows to construct a "well defined" ontological concept hierarchy. As an application of this approach, we presented some results of classification of clinical documents with maximally factorized common fields. We have shown that FCA can be useful method to classify and analyze various medical data by constructing concept hierarchy. From that concept hierarchy, we can acquire well-structured facts and knowledges in medical domain. (*Journal of Korean Society of Medical Informatics 12-1,31-43, 2006*)

*Key words:* Ontology, Formal Concept Analysis, Concept Hierarchy, Classification, Clinical Documents

논문투고일: 2005년 11월 25일, 심사완료일: 2006년 2월 2일

교신저자: 황석형, 충남 아산시 탕정면 선문대학교 컴퓨터정보학부 인문관 414호

전화: 041-530-2265, Fax: 041-530-2876, E-mail: shwang@sunmoon.ac.kr

\* 본 논문은 보건복지부 보건의료기술진흥사업의 지원(과제고유번호: A050909)에 의하여 연구되었음.

## I. 서론

계층적 개념구조(concept hierarchy)는 도메인 지식의 구조화와 요약을 제공하고, 필요한 정보의 수월한 접근을 제공하기 위해 사용되었으며<sup>1)</sup>, 또한 지식을 표현하는데 있어서 간결하면서도 효과적으로 개념간의 구조를 설명하기 위해 사용되고 있다. 이러한 개념구조의 대표적인 예로는, 사용자의 지식을 이용하여 정보에 접근하는 경로를 제공하는 ‘Yahoo!’의 디렉터리 서비스, 또는 지식을 나타내거나 WordNet과 같이 개념간의 is-a 관계(hyponym, hypernym)를 나타내기 위해 사용되는 트리 형태의 그래프 등이 있다.

그러나 지금까지 대부분의 계층적 개념구조는 특정 분야의 전문가나 지식 공학자들에 의해서, 많은 시간과 노력을 투입하여 작성되어 왔다. 또한, 오늘날과 같이 급변하는 환경에서 이미 구축된 계층적 개념구조는 지속적인 수정을 통해 새롭게 생성되는 개념 및 현상, 사물 등을 반영해야 할 필요가 있다. 따라서 본 논문에서는 일종의 개념적 클러스터링 기법으로서 널리 각광받기 시작한 형식개념분석기법(FCA: Formal Concept Analysis)을 이용하여 보다 수월하게 도메인 데이터로부터 개념계층구조를 구축하고 관리하는 방법을 제안하고, OO대병원에서 실제 사용하고 있는 각종 임상서식지들에 대하여 본 연구에서 제안한 형식개념분석기법을 적용하여, 임상서식지들을 분류하고 공통필드들을 추출하여 계층화함으로써 임상서식지 체계화에 본 연구 결과가 유용함을 제시하고자 한다.

본 논문은 다음과 같이 구성된다. 2장에서는 관련 연구의 소개와 형식개념분석기법을 이용한 계층적

개념구조의 구축방법에 대해 설명한다. 그리고 3장에서는 실제 적용사례로서, OO대학교병원의 임상서식지들을 대상으로 실시한 실험 결과를 보여주고, 4장에서는 결론과 향후 연구과제에 대해서 설명한다.

## II. 재료 및 방법

형식개념분석기법(Formal Concept Analysis)은 개념에 대한 속성을 정의하고 속성의 포함관계를 사용하여 개념적 계층구조를 생성하는 수학적 모델을 제공한다<sup>5)6)</sup>. 형식개념분석기법을 기반으로 하는 몇 가지 기존연구들<sup>2)3)</sup>로서는, Gu Tao는 온톨로지의 계층적 개념 트리의 구축을 돕기 위해 형식개념분석기법을 사용하는 방법을 제안하였다<sup>2)</sup>. 또한 형식개념분석기법을 사용하여 문서로부터 온톨로지의 개념 트리를 자동으로 구축하기 위해 문장에서의 동사와 목적어를 추출하여, 목적어를 개념으로, 동사를 개념의 속성으로 간주하여 형식개념분석기법에 적용하는 방법이 연구되었다<sup>3)</sup>. 또한, 객체지향 도메인 모델링의 중추적 역할을 하는 개념계층구조, 즉, 클래스계층구조를, 형식개념분석기법에서 제공하는 개념격자(Concept Lattice)를 기반으로 구축하기 위한 기법이 제안되었다<sup>7)</sup>.

이와 같은 기존의 연구들을 바탕으로, 본 논문에서는 실제 병원에서 사용되고 있는 임상서식지들을 대상으로, 형식개념분석기법을 이용하여 임상서식지들의 공통필드들을 추출하여 계층적 개념구조를 구축하는 방법을 제안하고, 관련 실험 결과를 살펴본다. 형식개념분석기법을 이용한 계층적 개념구조 구축 방법의 전체 구조는 Figure 1과 같이, (1)임상서식지들로부터 context table을 작성하고, (2)형식개념

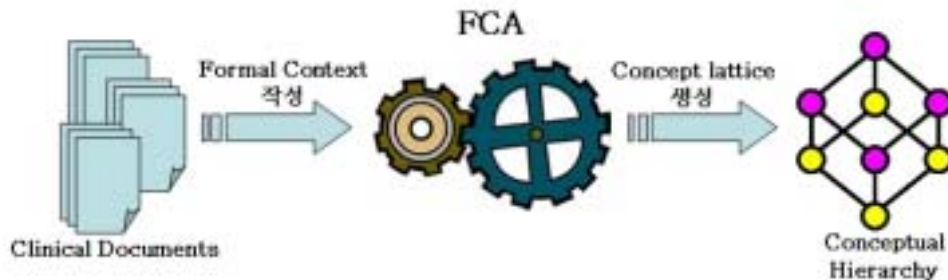


Figure 1. FCA-based approach to build conceptual hierarchy

분석기법을 적용하여, (3)개념격자구조(개념계층구조)를 구축하는 과정으로 이루어져 있다. 이와 같은 방법을 이용 함으로써 각 서식지에 공유될 수 있는 필드들을 최대한 추출하여 계층화함으로서 기존의 다종다양한 형태의 임상서식지들을 체계화 할 수 있다.

## 1. Formal Concept Analysis

형식개념분석기법(FCA: Formal Concept Analysis)은 개념(Concept)과 계층화된 개념구조에 관한 수학적 모델(개념격자, Concept Lattice)을 제공하여 제반 문제들에 대한 분석과 해결책을 모색하기 위한 응용수학분야의 기법으로서, 최근에는 의학 및 바이오인포메틱스, 사회과학, 소프트웨어공학, 온톨로지공학 등의 분야에 널리 적용되기 시작하고 있다<sup>4-6)</sup>. 형식개념분석기법의 기본이 되는 구조는 context이다. Context는 객체들(Objects)의 집합과 속성들(또는 특성들, 성질들, attributes)의 집합, 그리고 객체와 속성들 사이의 관계들(어떤 객체가 어떤 속성을 갖는가에 관한 관계들)로 구성된다. 형식개념분석기법을 보다 정확하게 이해하기 위하여 context의 정형화된 정의는 다음과 같다.

[정의 1] Formal context  $K=(G,M,I)$ 는 객체들(Objects)의 집합  $G$ 와 속성들(Attributes)의 집합  $M$ , 그리고  $G$ 와  $M$ 사이의 이항관계  $I \subseteq G \times M$ 로 구성된다. 즉,  $G$ 와  $M$ 의 원소들은 각각 해당 context의 객체들과 각 객체들이 가질 수 있는 속성들을 나타낸다. 또한, 어떤 객체  $g$ 가 속성  $m$ 을 가지고 있을 경우,  $gIm$  또는  $(g, m) \in I$ 로 나타내며,  $g$ 는  $m$ 을 갖는다는 것을 의미한다.

Formal context는 cross table형태로 나타낼 수 있으며, 해당 표의 행과 열의 헤더부분은 각각 context

를 구성하는 객체들과 속성들로 구성된다. 또한, table의 각 셀에 대해서는, 해당 셀에 관련된 객체와 속성이 이항관계  $I$ 를 만족할 경우에는  $\times$ 표시하고, 이외의 경우에는 빈 공간으로 남겨둔다. 예를 들면, 다음과 같은 동물들(Bat, Eagle, Monkey, Penguin, Shark)과 각 동물이 갖는 속성들로 구성되는 context table은 Table 1과 같다. 이와 같은 context로부터 개념들(concepts)을 추출할 수 있다. 각 개념들은  $(O, A)$ 와 같은 형태의 쌍(pair)으로 정의되며, 보다 정형적인 정의는 다음과 같다.

[정의 2] 임의의 context  $K=(G,M,I)$ 에 대하여,  $O \subseteq G$ ,  $A \subseteq M$ 일 때,  $\text{intent}(O) = A \wedge \text{extent}(A) = O$ 를 만족하는  $(O, A)$ 를 **개념**(formal concept)이라고 한다. 단,  $\text{intent}(O) := \{a \in M \mid \forall o \in O: (o, a) \in I\}$ ,  $\text{extent}(A) := \{o \in G \mid \forall a \in A: (o, a) \in I\}$ .

즉, formal concept  $(O, A)$ 는, intent와 extent로 구성되는 Galois Connection에 의해 규정되며,  $\text{intent}(O)$ 와  $\text{extent}(A)$ 에 의해  $O$ 의 모든 객체들이 공통적으로 갖는 속성들의 집합과  $A$ 의 속성들을 갖는 객체들의 집합을 각각 구할 수 있다.

Table 1의 임상서식지context에 대하여, 위의 정의2를 토대로 추출한 개념들은 Table 2와 같다. 이때 context로부터 추출된 개념들은 extent 또는 intent를 토대로 상위-하위개념관계(일종의 순서관계)를 정의할 수 있다.

[정의 3] 주어진 context  $K=(G,M,I)$ 의 임의의 개념  $(O1, A1)$ ,  $(O2, A2)$ 에 대하여, “상위-하위개념”관계  $(O1, A1) \leq (O2, A2)$ 는 일종의 반 순서관계(partial order relation)로서 다음과 같이 정의된다.

$$(O1, A1) \leq (O2, A2) \Leftrightarrow O1 \subseteq O2 (\Leftrightarrow A1 \supseteq A2).$$

이와 같은 상위-하위개념관계의 정의는, 임의의 개념은 외연적(extensional) 및 내포적(intensional)

Table 1. Context table for a vertebrate

	breathes in water	can fly	has beak	has hands	has skeleton	has wings	lives in water
Bat		×			×	×	
Eagle		×	×		×	×	
Monkey				×	×		
Penguin			×		×	×	×
Shark	×				×		×

인 관점에서 상위의 개념에 비해, 보다 적은 extension과 보다 큰 intension을 갖는다는 철학적인 개념 계층구조의 특성을 반영하고 있다. 특히, 주어진 context의 개념집합과 이들 요소들(개념들)간의 상위-하위개념관계는 일종의 반순서집합(poset: partially ordered set)에 해당하며, 이를 개념격자(Concept Lattice 또는 Galois Lattice)라고 부른다.

## 2. Concept Lattice를 이용한 개념계층의 구조화

개념격자는 Hasse Diagram을 사용하여 계층적인 개념구조형태로 가시화할 수 있다. Hasse Diagram은 정점들과 링크들로 구성되며, 각 정점은 개념격자의 개념들에 대응하며, 각 정점에는 2종류에 레이블(Extension과 Intension)이 각각 하단과 상단에 표시된다. 각 개념들 사이를 연결하는 링크들은 개념들 사이의 “상위-하위개념 관계”를 나타낸다. 특히, 개념격자를 나타낸 Hasse Diagram에서는 상위부분의 정점들은 보다 추상적이고 일반화된 개념들을, 반면에 하위부분에는 보다 구체화된 개념들이 위치하게 된다.

FCA의 제반정의들을 토대로 주어진 context로부

터 개념들을 추출하여 상위-하위개념관계를 파악하여 개념격자구조를 생성하기 위한 알고리즘은 다음과 같다.

Table 1의 context에서 추출한 개념들(Table 2)에 대하여 Hasse Diagram을 사용하여 개념격자를 표기하면 Figure 3과 같다. 개념격자를 나타낸 Hasse Diagram에서는, 각 개념들과 이들 사이의 상하위관계가 정점과 링크에 의해 각각 표시되며, 특히, 개념들 간의 링크에 의해 만들어지는 경로에 의해 상위개념으로부터 하위개념으로 속성들이 상속되며, 하위개념으로부터 상위개념으로 해당 객체들이 전파된다. 예를 들면, Figure 3에서 “Shark”는 고유의 속성으로서 “breathes in water”를 갖고 있으며, 더불어서 상위개념들로부터 “has skeleton”과 “lives in water”와 같은 속성들을 상속받는다. 한편, “has wings” 속성을 갖는 객체로서는 “Bat”, “Eagle”, “Penguin”이 됨을 알 수 있다. 또 개념격자구조를 사용하여 개념들 간의 속성에 관한 추론을 수월하게 할 수 있다. 예를 들면 Figure 3에서 “breathes in water”이면 “lives in water”이고 “has skeleton”이다. 또한 “can fly”인 것은 “has wings”이고 “has skeleton”이다와 같은 추론을 할 수 있다. 이와 같은 형식개념분석기법을 사용함으로써, 주어진 문제영역

---

```

// INPUT : a formal context C = (O, A, R)
// OUTPUT : Concept Lattice L := (B(C), E≤)

for all o∈O do
    B(C) ← B(C) ∪ (extent(intent(o)), intent(o));
end for
for all c∈B(C) do
    for all o∈(O - extent(c)) do
        X ← extent(c) ∪ {o};
        if (extent(intent(X)), intent(X)) ∉ B(C) then
            B(C) ← B(C) ∪ (extent(intent(X)), intent(X));
        end if
    end for
end for
for all c1∈B(C) do
    for all c2∈B(C) - {c1} do
        if ((c1≤c2) ∧ (∄c3∈B(C) - {c1, c2}[(c1≤c3) ∧ (c3≤c2)])) then E≤ ← E≤ ∪ {(c1, c2)};
        end if
    end for
end for

```

---

Figure 2. Algorithm : generate concepts and build concept lattice

Table 2. All formal concepts of table 1

Concepts	Extensions	Intensions
C1	$\emptyset$	{breathes in water, can fly, has beak, has hands, has skeleton, has wings, lives in water}
C2	{Eagle}	{can fly, has beak, has skeleton, has wings}
C3	{Monkey}	{has hands, has skeleton}
C4	{Penguin}	{has beak, has skeleton, has wings, lives in water}
C5	{Shark}	{breathes in water, has skeleton, lives in water}
C6	{Bat, Eagle}	{can fly, has skeleton, has wings}
C7	{Eagle, Penguin}	{has beak, has skeleton, has wings}
C8	{Penguin, Shark}	{has skeleton, lives in water}
C9	{Bat, Eagle, Penguin}	{has skeleton, has wings}
C10	{Bat, Eagle, Monkey, Penguin, Shark}	{has skeleton}

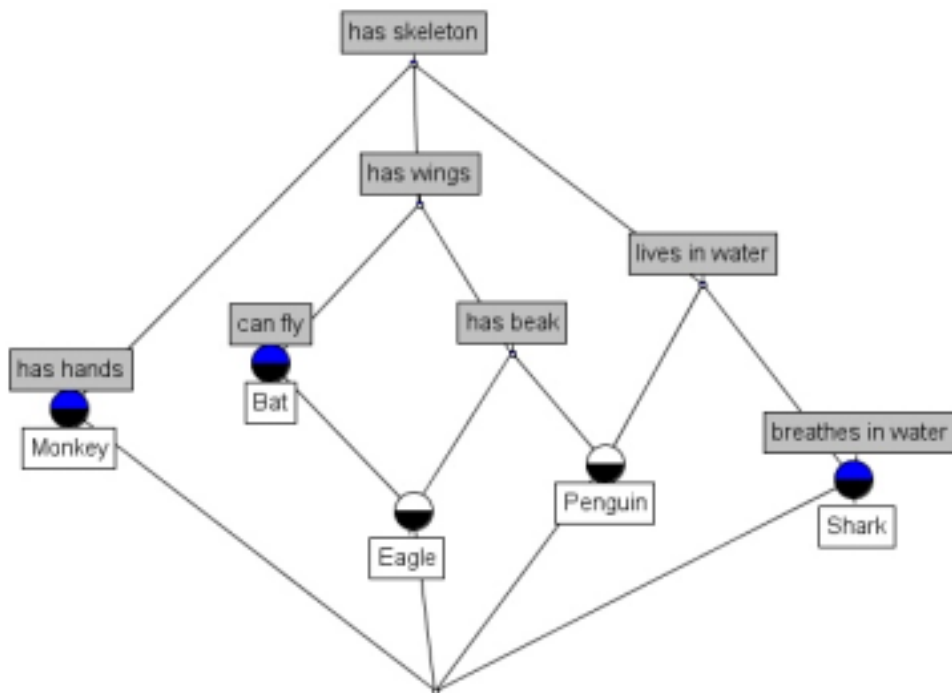


Figure 3. Concept lattice for table 1

의 객체들과 이들이 갖는 속성들을 context 형태로 파악하여, 개념을 추출하며 개념격자 형태로 나타냄으로써, 도메인 내의 개념들을 공통속성이 추출된 형태로 분류할 수 있는 계층적 개념구조를 수월하게 구축할 수 있다.

OO대병원에서 실제로 사용되고 있는 임상서식지

들을 대상으로, 문서집합들을 구조화하여 필요한 정보를 수월하게 획득하고 관련 지식을 간결하면서도 효과적으로 표현하기 위한 계층적 개념구조를 구축하기 위하여 형식개념분석기법(Formal Concept Analysis)을 적용한 문서체계화 실험을 수행하였다. 본 논문에서는 이와 같은 일련의 형식개념분석을 적

용한 실험을 수행하기 위해 ConExp프로그램<sup>8)</sup>을 사용했다.

### III. 결 과

#### 1. 임상서식지들로부터 context table을 작성

00대학병원에서 실제로 사용하고 있는 임상서식지들 중에서 임의로 추출한 8개의 서식지들(간호정보기록, ICU간호정보기록, 처방진행기록, 입퇴원기록, 응급의료센터 임상기록, 응급의료센터 기록지, 집중치료실 기록, 간호기록)을 대상으로 Table 3과 같은 context table을 작성하였다. 즉, Table 3의 객체는 간호정보기록, ICU간호정보기록, 처방진행기록, 입퇴원기록, 응급의료센터 임상기록, 응급의료센터 기록지, 집중치료실 기록, 간호기록 등과 같은 8종류의 임상서식지들이며, 속성은 8종류의 임상서식지들에 포함되어 있는 84개의 필드들로 구성된다. 또한, 각 임상서식지들과 필드들 사이의 관계들(예를 들면, 간호기록 서식지는 병상, 병동, 시행(I), 기록(S.O.A.P.E) 등의 필드들을 포함하고 있음을 나타내는 이항관계)을 해당 셀에 X표시함으로써 임상서식지들로부터 context table을 작성하였다.

단, 위에서 ‘기록(S.O.A.P.E)’이라고만 표기되어 있는 부분은, 주관적 자료(subjective data), 객관적 자료(objective data), 사정(assessment), 계획(planning), 평가(evaluation)의 앞 글자만을 떼어서 만든 용어로서, 문제 중심의 간호기록에서 비롯된 것으로 많은 기관에서 이용되고 있다. 여기서, 주관적 자료는 대상자 자신이 경험하고 느끼고, 보고, 듣고, 혹은 생각하고 있는 것들을 말로 나타낸 것이고 객관적 자료는 활력징후, 의료요원의 감각을 사용한 관찰내용, 검사와 엑스선 촬영결과 그리고 검사와 치료 등에 대한 대상자의 반응이며, 사정단계에서 관찰자는 주관적, 객관적 자료를 이용하여 해석하고 결론을 내리게 되며 이 단계에서 간호사의 지식과 논리적인 사고가 요구된다.

따라서, S.O.A.P.E와 같이 주관적이고 경험과 생각 등을 토대로 하는 사항들에 대해서는, 본 논문에서 제안하고 있는 One-Valued Context형식으로 정

형화하는 것(A라는 임상서식에는 “가”, “나”, “다” 등과 같은 항목들이 포함/불포함되어 있다는 정보를 binary형태의 context table로 만드는 것)은 다소 무리가 있을 수도 있으나, 현재 필자들이 개발 중인 Many-valued context형식(A라는 임상서식에는 “가”, “나”, “다” 등과 같은 항목들이 포함되어 있고, 각각의 항목에는 “X”, “Y”, “Z”라는 값이 들어있다는 정보를 table형태로 만든 것)을 이용할 경우, S.O.A.P.E항목과 같은 유형의 문제는 해결될 수 있다.

특히, Many-valued context에 대하여, FCA기법을 적용시키기 위해서는 다시 One-valued context로 변환하는 작업이 필수적이며, 이러한 변환작업에는 의료분야의 의사, 간호사, 의무기록사 등, 해당분야 전문가들의 지식과 경험, 생각, 그리고 이를 체계화한 해당도메인의 온톨로지가 절대적으로 요구된다. 따라서, 이와 같은 사항들을 전부 고려하여 해당 변환작업을 정형화하기 위하여, 현재 관련연구를 진행중에 있으므로 본 논문에서는 상술하지 않는다.

#### 2. Concept 추출

본 논문에서 제안한 형식개념분석기법을 토대로 ConExp<sup>8)</sup>를 사용하여 임상서식지 Context Table로부터 Table 4와 같은 15개의 개념<sup>1)</sup>들이 추출되었다. 각각의 개념들은 Extension와 Intension, 즉, 해당 개념을 구성하는 객체들과 속성들로 구성되어 있다. 단, 여기서 객체는 임상서식지들의 집합이고, 속성은 임상서식지들이 공통으로 포함하고 있는 필드들의 집합으로 구성되어 있다. 예를 들어, Table 4의 C7개념은 “집중치료실기록” 서식지를 나타내는 개념으로서 “집중치료실 입원일수, 배액, 간호활동, 검사수치, 교환일, 피부, 순환계” 등과 같은 25개의 필드들을 가지고 있다. 또한 C9개념은 “간호기록” 서식지와 “집중치료실기록” 서식지의 공통개념으로서 “시행(I), 기록(S.O.A.P.E), 병동, 병상, 등록번호” 등 8개의 필드를 공통으로 가지고 있다. 이와 같이

1) 여기서 언급하고 있는 “개념”, 즉 “Concept”은, 임상서식지들과 공통항목들의 쌍으로 구성되므로, 앞으로 논문의 본문 내에서는 “임상서식지들을 구성하는 공통항목요소들”을 지칭한다.



FCA를 이용한 임상서식지의 분류와 체계화

Table 4. All formal concepts of table 3

Concepts	Extensions	Intensions
C1	∅	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 입원경로, 정보제공자, 입원경력, 가족력, 환자요구사항, 최근투약내용, 입원하게된 동기, 예방접종<소아>, 알러지, 영양대사, 의치, 심장혈관관계/순환상태, 수면습관, 신경계/운동, 배변배뇨, 귀중품, 투약기간, 약물복용, 작성일, 발생시기, 수술명, 실시내역, 간호처치, 환자전화번호, 환자주소, 주진단명, 치료결과, 퇴원형태, 의무기록사 서명, 재원일수, 수술및 치료명, 입력완료, 부검, 수간호사서명, 기록완성일자, 부진단명, 입원과, 내원시교육, 내원일시, 내원수단, 치료내용, 전공의/시간, 진찰시간, 입원장 발부시간, 주증상, 서명, 응급치료후 결과, 사고구분, 사고발생기간, 고지과, 사고발생경위, 인공호흡기, 산소포화도, 입원실, 투약내용, 시술자, 인계사항, 동공반사, 부착물, 검사, 내원경로, 시행(I), 기록(S.O.A.P.E), position, suction, 전공의, 집중치료실 입원일수, 삼입일, 순환계, 위장계, 피부, 드레싱, 배액, 비뇨기계, 중증도, 간호활동, 검사수치, 교환일}
C2	{간호정보기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 입원경로, 정보제공자, 입원경력, 가족력, 환자요구사항, 최근투약내용, 입원하게된 동기, 예방접종<소아>, 알러지, 영양대사, 의치, 심장혈관관계/순환상태, 수면습관, 신경계/운동, 배변배뇨, 귀중품}
C3	{ICU간호정보기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 입원경로, 정보제공자, 입원경력, 가족력, 투약기간, 약물복용, 작성일, 발생시기, 수술명}
C4	{입퇴원기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 환자전화번호, 환자주소, 주진단명, 치료결과, 퇴원형태, 의무기록사 서명, 재원일수, 수술 및 치료명, 입력완료, 부검, 수간호사 서명, 기록완성일자, 부진단명}
C5	{응급의료센터입상기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 환자전화번호, 환자주소, 주진단명, 입원과, 내원시교육, 내원일시, 내원수단, 치료내용, 전공의/시간, 진찰시간, 입원장 발부시간, 주증상, 서명, 응급치료후 결과, 사고구분, 사고발생기간, 고지과, 사고발생경위}
C6	{응급의료센터기록지}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 인공호흡기, 산소포화도, 입원과, 내원시교육, 내원일시, 내원수단, 입원실, 투약내용, 시술자, 인계사항, 동공반사, 부착물, 검사, 내원경로}
C7	{집중치료실기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 인공호흡기, 산소포화도, position, suction, 전공의, 집중치료실 입원일수, 삼입일, 순환계, 위장계, 피부, 드레싱, 배액, 비뇨기계, 중증도, 간호활동, 검사수치, 교환일, 시행(I), 기록(S.O.A.P.E)}
C8	{처방진행기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 실시내역, 간호처치}
C9	{간호기록, 집중치료실기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 시행(I), 기록(S.O.A.P.E)}
C10	{간호정보기록, ICU간호정보기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 입원경로, 정보제공자, 입원경력, 가족력}
C11	{입퇴원기록, 응급의료센터입상기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동, 환자전화번호, 환자주소, 주진단명}
C12	{응급의료센터입상기록, 응급의료센터기록지}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 입원과, 내원시교육, 내원일시, 내원수단}
C13	{응급의료센터기록지, 집중치료실기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 인공호흡기, 산소포화도}
C14	{간호정보기록, ICU간호정보기록, 처방진행기록, 입퇴원기록, 응급의료센터입상기록, 간호기록, 집중치료실기록}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호, 병상, 병동}
C15	{간호정보기록, ICU간호정보기록, 처방진행기록, 입퇴원기록, 응급의료센터입상기록, 간호기록, 집중치료실기록, 응급의료센터기록지}	{환자성별/나이, 환자이름, 주민등록번호, 등록번호}



context로부터 개념들을 추출함으로써 각 서식들이 갖는 고유항목 뿐만 아니라 각 서식지간의 공통항목들을 수월하게 추출 할 수 있다.

### 3. 개념격자를 이용한 개념계층구조화

이와 같이 추출된 각 개념들을 토대로, 개념들 사이의 상위-하위개념관계를 파악하여 계층화된 형태의 개념격자구조로 표현한다(Fig. 4). Figure 4의 개념격자에서 C7는 “집중치료실기록” 서식지를 나타내는 개념으로서 “전공의, 순환계, 위장계, 피부, 드레싱, 중증도, 간호활동”등과 같은 고유의 필드들 뿐만 아니라 C13개념의 필드들(“인공호흡기, 산소포화도”)과, C9개념의 필드들(“시행(I), 기록(S.O.A.P.E)”), C14개념의 필드들(“병상, 병동”), C15개념의 필드들(“등록번호, 환자이름, 주민등록번호, 환자성별/나이”)을 포함하고 있다. 또한 C9는 “간호기록, 집중치료실기록” 서식지를 나타내는 개념으로서 “시행(I), 기록(S.O.A.P.E)”과 같은 고유의 필드들 뿐만 아니라 상위의 C14, C15개념의 필드들(“병동”, “병상”, “환자성별/나이”, “환자이름”, “주민등록번호”, “등록번호”)을 공통적으로 가지고 있음을 나타내고 있다.

한편, C13과 같은 개념의 경우, 해당하는 서식지는 없으나 “응급의료센터 기록지” 및 “집중치료기록”서식지들에 공유될 수 있는 공통적인 필드들(“인공호흡기, 산소포화도”)을 고유의 필드로 보유하고 있는 개념이다. 이외에도 “간호정보기록”과 “ICU간호정보기록”서식지들에 공유될 수 있는 공통적인 필드들(“입원경로, 정보제공자, 입원경력, 가족력”)을 추출한 개념으로서 C10이 발견되었으며, “입퇴원기록”과 “응급의료센터 임상기록”서식지들에서 공유될 수 있는 공통적인 필드들(“환자전화번호, 환자주소, 주민단명”)을 추출한 개념으로서 C11을 발견할 수 있었다. 특히, C15는 전체 8개 서식지들(“간호정보기록, ICU간호정보기록, 처방진행기록, 입퇴원기록, 응급의료센터 임상기록, 응급의료센터 기록지, 집중치료실 기록, 간호기록”)에서 공유할 수 있는 필드들(“환자 성별/나이, 환자이름, 주민등록번호, 등록번호”)을 발견하여 추출된 개념으로서, 이들 필드들을 실험대상서식지 전체에서 공유할 수 있음을

알 수 있다.

Figure 5에서는 Figure 4에 나타난 개념격자구조상에 있는 각 개념들에 대한 서식지개수와 비율을 나타내고 있다. 예를 들어 C9개념에 속한 서식지는 전체 8개의 서식지의 25%에 해당하는 2개서식지(“간호기록”, “집중치료실 기록”)를 나타내고 있다. 또 C15개념에는 8개의 전체서식지가 모두 속해 있으므로 100%의 서식지 비율을 확인 할 수 있다. Figure 5의 C9~C13으로부터 공통의 필드들은 전체 서식지의 25%에 해당하는 서식지로부터 추출 될 수 있음을 알 수 있다. 또한, C14의 공통필드(병동, 병상)는 전체서식지의 88%에 해당되는 서식지에서 공유 할 수 있으며, C15는 전체서식지의 100%인 모든 서식지에서 공통의 필드들을 공통적으로 사용할 수 있음을 알 수 있다.

한편, Figure 6에서는 Figure 4에 나타난 개념격자구조상에 있는 각 개념들에 대한 필드의 개수와 비율을 나타내고 있다. 예를 들어 C9개념은 전체 84개의 필드 중 10%에 해당하는 8개의 필드들(“시행(I), 기록(S.O.A.P.E), 병상, 병동, 환자성별/나이, 환자이름, 주민등록번호, 등록번호”)을 가진 개념으로서, “간호기록” 및 “집중치료실기록”서식지에 공통적으로 사용할 수 있음을 나타내고 있다. 또한, C15개념은 전체 84개 필드들 중 5%에 해당하는 4개의 필드들(“환자성별/나이, 환자이름, 주민등록번호, 등록번호”)을 가지고 있는 개념으로서, 실험대상이 된 전체 8개 서식지에서 공통적으로 사용될 수 있음을 보여주고 있다. 특히, Figure 6의 C9~C15로부터, 실험대상이 되고 있는 8개 서식지 중에서 임의의 2개 문서에 공통적으로 공유되고 있는 필드는 전체의 5~12%에 해당함을 알 수 있었다.

## IV. 고 찰

본 논문에서는 형식개념분석기법(FCA: Formal Concept Analysis)을 기반으로 하여 계층적 개념구조를 구축하기위한 기법을 제안하였다. 또한, 의료분야의 문서체계화를 위하여, OO대학교병원의 일부 임상서식지들에 대하여 형식개념분석기법을 적용하여 계층화된 문서체계를 구축 하기 위한 실험을 수행하였다. 즉, 본 실험에서는 실험대상문서들과 각

FCA를 이용한 임상서식지의 분류와 체계화

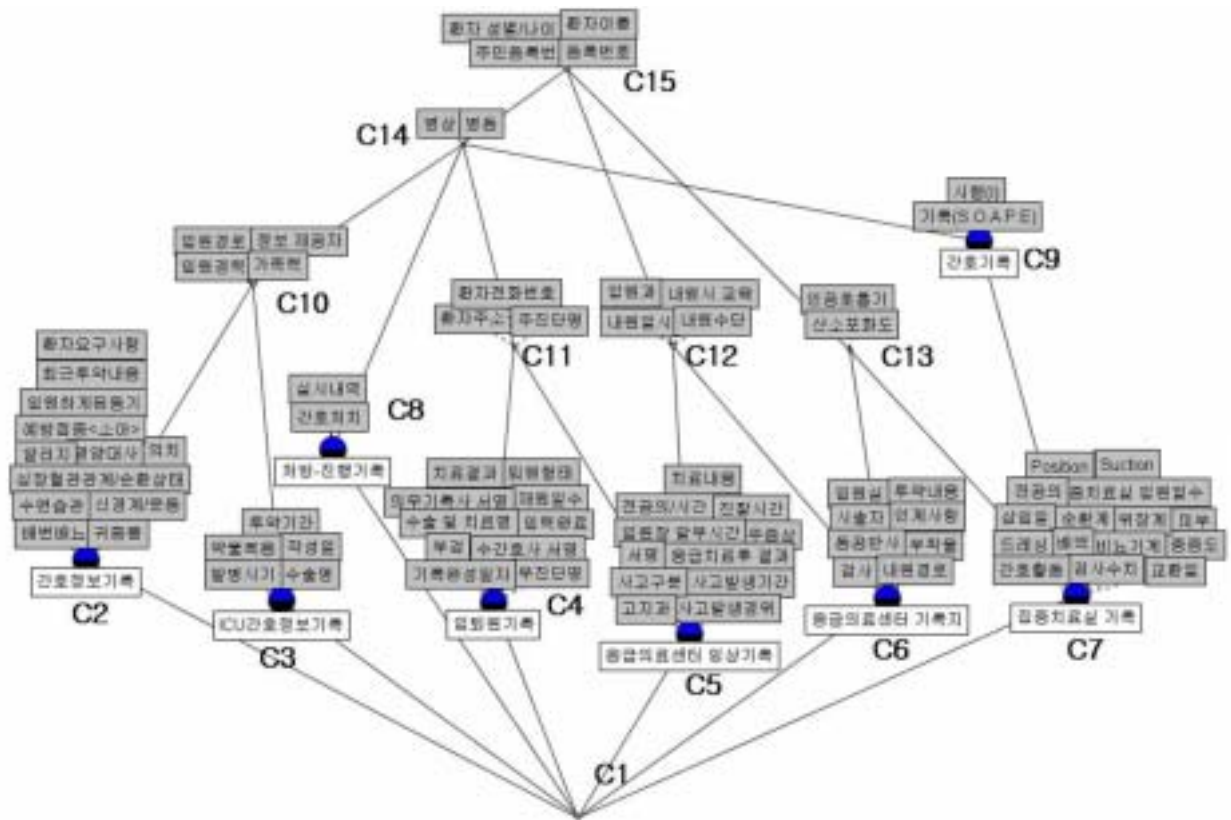


Figure 4. Concept lattice for table 3

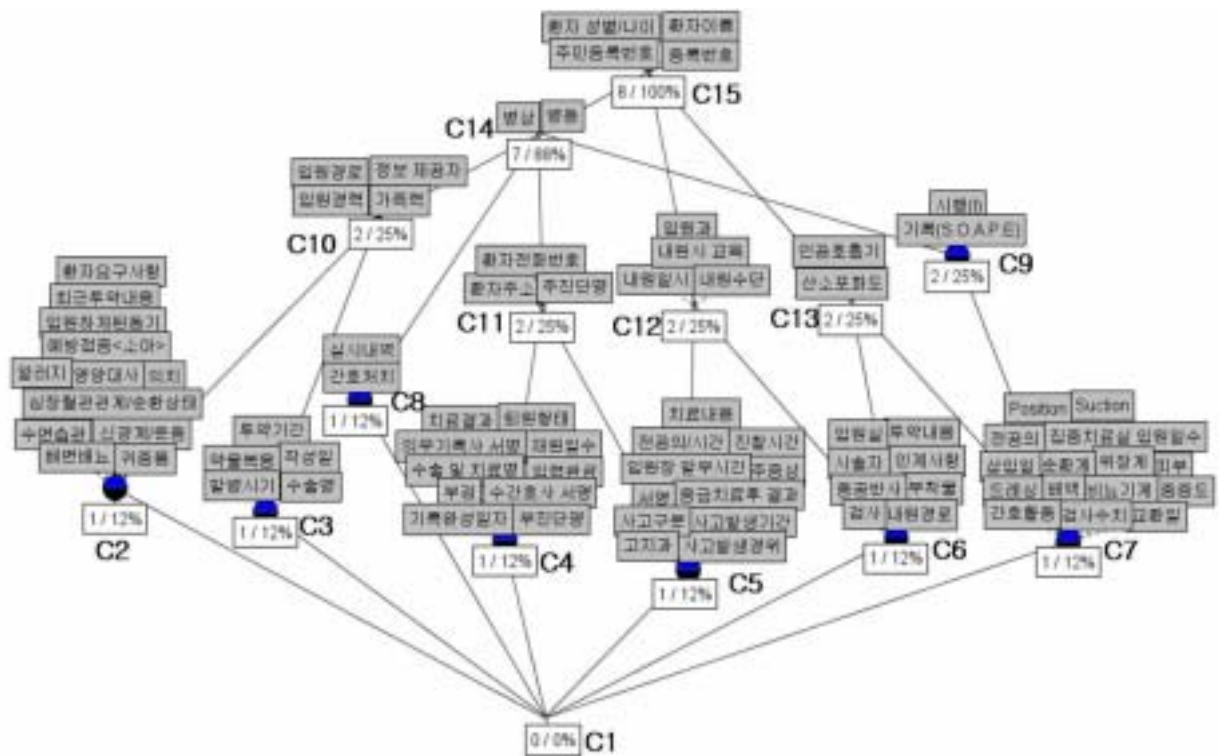


Figure 5. Concept lattice with additional information for number of documents

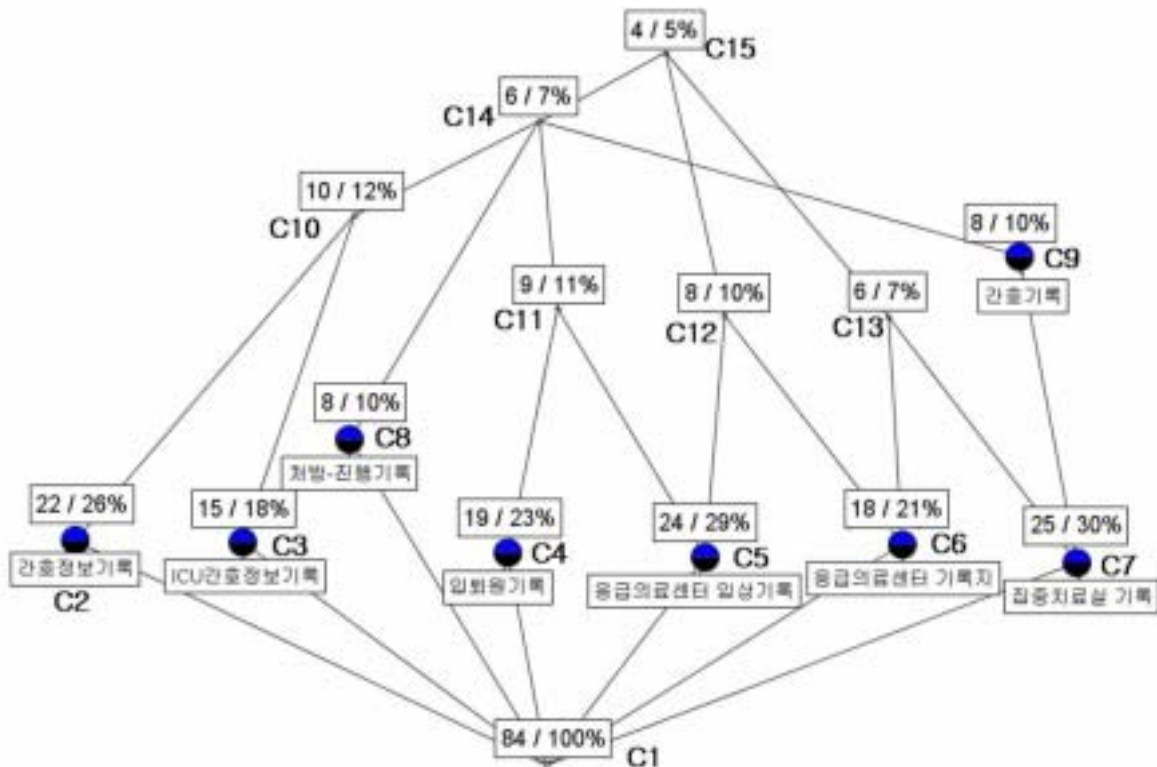


Figure 6. Concept lattice with additional information for number of field

문서들을 구성하고 있는 필드들 간의 관계를 context table형태로 파악하여 공통적인 속성들을 갖는 개념들을 추출하여 계층적 개념구조형태로 추상화함으로써, 각 문서들이 공유할 수 있는 공통 필드들을 최대한 추출한 형태의 문서 체계화가 가능하다.

본 연구에서 실시한 실험결과, OO대병원에서 사용되고 있는 임상서식지들 중에서 임의로 추출한 8종류의 임상서식지에 대하여 15개의 개념을 추출할 수 있었고, 7개의 공통부분을 발견해 낼 수 있었다. 본 실험결과를 토대로 체계화된 임상서식지 계층구조를 구축 할 수 있고, 이와 같은 계층구조는, 현재 사용되고 있는 모든 임상서식지들에 대한 모듈화 및 체계적인 문서관리, 그리고 관련정보의 효율적인 검색 등을 원활하게 수행할 수 있는 기반이 될 수 있다. 단, 본 연구에서는 실험대상 서식지의 종류가 8개에 불과하므로 보다 유효한 결과를 얻기 위해서는 실험대상을 확대할 필요가 있다.

특히, 본 논문에서는 실제 임상에서 사용되고 있는 서식지들을 토대로, Bottom-up방식으로 임상서식지들을 분류하여 체계화하기 위한 정형화된 기법

을 제안하였으며, 본 기법의 적용사례로서, OO대병원에서 실제로 사용되고 있는 임상서식지들 중에서 임의로 추출한 8종류의 서식지들(간호정보기록, ICU간호정보기록, 처방진행기록, 입퇴원기록, 응급의료센터 임상기록, 응급의료센터 기록지, 집중치료실 기록, 간호기록)을 분류하여 체계화시킨 실험결과를 서술하고 있다. 특히, 본 실험결과(개념격자구조로 체계화된 임상서식지 계층구조)로부터, 8종류의 서식지들 사이에 공유될 수 있는 공통항목들을 본 논문에서 제안한 기법에 의해 수월하게 자동적으로 추출할 수 있음을 서술하였다.

기존의 유사한 연구(임상서식지 분류와 관련된 연구)<sup>10-14)</sup>들에서는 의료분야전문가(의사, 간호사, 의무기록사 등)가 국내의료정보 표준화를 위하여 다양한 의무기록양식을 수집하고 각계의 의견을 수렴하여 수작업에 의해 공통항목들을 정리하였다. 그러나, 기존의 연구결과들 중에서는 문서체계화 결과가 이상적인 결과라기보다는 의료분야의 현실을 반영한 결과로서 실험대상병원의 서식항목을 가지고 표준안으로 일반화시키기에 무리가 따를 수 있다는 문

제점을 언급하고 있다<sup>12)</sup>. 이와 같은 기존연구결과들에 대하여, 본 논문에서 제안한 FCA기법을 적용할 경우, 기존 유사연구의 타당성과 유용성을 입증할 수 있을 뿐만 아니라, 더 나아가서는 수작업에 의해 발견하지 못했던 공통항목들을 수월하게 추출할 수 있다. 따라서, 본 연구결과는 기존의 의무기록양식의 체계화를 보완할 수 있으며, 향후의 EHR과 같은 의료정보 표준화를 기반으로 하는 전산화에 기여할 수 있을 것으로 생각한다.

FCA기반의 임상서식지의 분류 및 체계화기법은, 의료분야 전문가들(의사, 간호사, 의무기록사 등)의 지식을 토대로, 실제 병원에서 사용되고 있는 다종 다양한 서식지들에 대하여 반복적이고 점증적으로 적용시킴으로써, 기존의 표준화작업에 의해 정형화된 서식지체계<sup>10-14)</sup>를 보완할 수 있는 가능성을 제공하고 있다. 따라서, 이와같은 상호보완효과를 극대화시키고 타당성과 신뢰성을 검증하기 위해서는 국내 외의 다양한 병원에서 사용되고 있는 임상서식지에 대한 다양한 실험과 더불어, EMR 또는 EHR분야에 까지 확대해서 적용해 볼 필요가 있으며, 이를 위한 후속연구가 진행중에 있다.

따라서, 향후 연구과제로서는, 본 논문에서 제안한 의료문서 체계화 기법을 토대로 하여, 병원내의 각 진료과 또는 각 병원들을 대상으로 보다 광범위한 형태의 문서체계화가 필요하다. 이와 같이 의료분야의 문서체계화를 위해서는 병원에서 실제로 사용하고 있는 구체적인 상세한 임상서식지들이 모두 망라 되어야하며, 보다 정확하고 유용한 결과를 얻기 위해서는 의료분야 전문가의 도움이 절대적으로 필요하다.

### 참고문헌

1. Lawrie D, Croft WB. Discovering and comparing topic hierarchies. Proceedings of RIAO2000 conference:2000 April 12-14: Paris, France.
2. Gu T. Using formal concept analysis for ontology structuring and building [dissertation]. Singapore:Nanyang Technological University: 2003
3. Cimiano P, Staab S, Tane J. Automatic

- acquisition of taxonomies from Text-FCA meets NLP. Proceedings of the ECML/PKDD Workshop on Adaptive Text Extraction and Mining:2003 September 22; Cavtat-Dubrovnik, Croatia.
4. Davey BA, Priestley HA. Introduction to lattices and order. 2nd ed. UK:Cambridge University Press:2002. pp.65-84.
5. Ganter B, Wille R. Formal concept analysis. 1st ed. Heidelberg:Springer-Verlag:1999. pp.17-62.
6. Carpineto C, Romano G. Concept data analysis. 1st ed. UK:Wiley:2004. pp.21-23.
7. Hwang SH, Kim HG, Yang HS. A FCA-based ontology construction for the design of class hierarchy. Lecture Notes in Computer Science 2005;3482(3):827-835.
8. Available at :  
<http://sourceforge.net/projects/conexp>.  
Accessed October 1, 2005
9. Hwang SH, Kang YK, Kim HG, Kim MK. A FCA-based conceptualization of clinical documents. Journal of Korea society of Medical Informatics 2005;11(Suppl 1):53-56.
10. Choi JW, Hong JH, Kim ON, Seo JS, Yoo YS, Boo YK, et al. Review of core elements used in medical records and definitions for medical informatics standard. Journal of Korea society of Medical Informatics 1997;3(1):233-238.
11. Cho HI, Choi JW, Kim KD, Nam SM, Hong JH, Kim ON, et al. Review of core elements used in medical records and their definitions for medical informatics standard. Journal of Korea society of Medical Informatics 1997;3(2):91-98.
12. Lee JH, Lee HJ, Chae YM, Hong JH. Standardization study for discharge abstract data. Journal of Korea society of Medical Informatics 1998;4(1):15-28.
13. Park HA, Cho IS, Kim KD, Kim SH, Park JS, Lee YS, et al. Analysis and standardization of nursing record forms for nursing informatics standard. Journal of Korea society of Medical

Informatics 1998;4(2):69-79.  
14. Park HA, Cho IS, Kim KD, Park JS, Yoo KS,  
Yoon SJ, et al. Standardization of nursing

documents for special nursing units. Journal of  
Korea Society of Medical Informatics 2000:  
6(3):31-38.

