

## 해양GIS XML 응용스키마 결합방법 연구

### A study on integration of XML application schema for MGIS

오세웅\* · 박계각\*\* · 박종민\* · 서상현\*

Se-Woong Oh\*, Gyei-Kark Park\*\*, Jong-Min Park\* and Sang-Hyun Suh\*

\* 한국해양연구원 해양시스템안전연구소 해양안전 · 방제기술연구부

\*\* 목포해양대학교 해상운송시스템 학부

#### 요 약

분산되어 있거나 이기종의 데이터베이스 통합을 위해 많은 연구가 수행되었으며, 특히 데이터베이스 통합을 위해 데이터베이스의 스키마 매칭이 이슈가 되고 있다. 본 연구에서는 해양GIS 데이터베이스 통합을 위해 해양GIS 데이터베이스 구조를 XML 응용스키마로 변환하고 유사도 측정법과 relaxation labeling 기법을 이용하여 매칭 방법을 연구하였다. XML 응용스키마의 항목 간 유사도 측정은 각 항목이 가지고 있는 범주와 속성으로 계산하며, 각 항목이 계층간 의미적 연계관계를 적용하기 위하여 relaxation labeling 기법을 이용하여 초기 유사도 측정값에 반복적으로 적용하여 유사도 측정값의 변이를 측정하였다. 본 연구에서는 해양GIS 메타데이터 응용 스키마의 예시를 들어 알고리즘을 통한 매칭을 시도하였고, 실험 결과 알고리즘에 따른 초기 계산 값의 변화를 확인하였다.

키워드 : 해양GIS, XML, 응용스키마, 통합, 메타데이터

#### Abstract

Information integration for distributed and heterogeneous data sources is still an open challenging, and schema matching is critical in this process. This paper presents and approach to automatic elements matching between XML application schemas using similarity measure and relaxation labeling. The semantic modeling of XML application schema has also been presented. The similarity measure method considers element categories and their properties. In an effort to achieve an optimal matching, contextual constraints are used in the relaxation labeling method. Based on the semantic modeling of XML application schemas, the compatible constraint coefficients are devised in terms of the structures and semantic relationships as defined in the semantic model. To examine the effectiveness of the proposed methods, an algorithm for XML schema matching has been developed, and corresponding computational experiments show changes of calculated values.

Key Words : MGIS, XML, Application schema, Integration, Metadata

### 1. 서 론

연안 및 해양환경 등의 중요성이 부각되면서 해양공간의 통합적이고 효율적인 계획과 개발, 활용, 보존을 위한 지리정보 및 서비스 기술의 요구가 증대되고 있다. 지리정보 분야 국제 표준 기구에서는 상호 표준의 연계를 시도하고 있으며 미국, 캐나다, 호주, 일본, 영국 등의 국가들은 국제기구와 연계하여 국가 해양지리정보 표준 사양의 상호 연계성 및 호환성을 높이기 위해 다양한 방안을 모색 중에 있다.

이에 따라 해양지리정보 관련 기관에서는 각 기관마다 독립적으로 생산되어 분산 혹은 이기종의 해양지리정보 데이터베이스를 통합하는 연구가 시도된 바 있다. 해양지리정보 통합 절차 중 데이터베이스 스키마의 매칭(Matching)은

데이터 통합에 있어서 핵심적인 문제로서, 통합하고자 하는 두 스키마를 비교하여 항목 간 매핑방법을 제공하였다. 현재, 해양지리정보 데이터베이스 통합 방법은 수동으로 조작하거나, 많은 시간이 소요되는 작업으로서 자동화에 대한 필요성이 제기되어 왔다. 본 연구에서는 해양지리정보 통합을 위한 방법으로 클래스 항목 간 유사도 측정법과 클래스 항목 간 의미적 연계관계 고려를 위한 relaxation labeling 기법을 이용한 스키마 통합 방법을 개발하였다. 본 개발 방법의 사례 연구로서 해양지리정보의 공급 및 관리를 위해 사용되고 있는 해양지리정보 메타데이터 스키마를 예제로 적용하였다.

### 2. 해양지리정보의 통합

해양지리정보란 해양공간정보와 속성정보를 가진 데이터로서 국립해양조사원에서 생산하는 전자해도, 수치해도, 어업정보도, 해양수산 관련 기관에서 생산하는 각 주제도를 포함한다. 특히 해양지리정보는 수로 및 해저지형 측량사업,

접수일자 : 2008년 11월 1일

완료일자 : 2009년 3월 10일

본 연구는 기초기술연구회의 “해양 유출사고 대응지원 시스템 구축(PG47272)”, 국립해양조사원의 “다이나믹 전자해도 개발(PGS2070)” 연구의 지원을 받아 수행되었습니다.

전자해도 및 수치지도 사업, 해안선 조사, 측량 및 DB 구축 관련 사업, 연안관리정보 구축 사업, 해양기초지리정보 구축 관련 사업 등을 통해 생산되고 있다. 해양수산 관련기관에서 생산하는 주제도는 해안선, 수심, 항로표지, 항만시설과 같이 많은 공통부분을 포함하고 있으나, 해양지리정보 구축 사업이 독립적으로 시행됨에 따라 중복 자료가 생산되고 있다. GIS 프로젝트 예산중 80% 이상이 데이터 생산에 소요됨을 감안할 때, 예산 절감과 효율성을 위해 해양지리정보의 통합이 필요하다.

해양지리정보는 동적인 매질의 특성으로 인해 정보의 유지관리가 매우 중요하며 따라서 높은 유지관리 비용이 발생되고 있다. 해양지리정보 데이터베이스 구조가 통합될 경우 단일의 유지관리가 가능하며 비용절감 효과가 있다. 이에 본 연구에서는 다음과 같이 해양지리정보 데이터베이스 통합 방법을 설계하였다. 확장 가능한 마크업 언어인 XML(eXtensible Markup Language) 기술의 도입으로 자료의 저장, 모델링, 전송 기술이 발전하여, 데이터베이스 분야에 있어서도 XML Schema 기술을 이용한 데이터베이스 구조 설계가 이루어지고 있으며, 오라클, MS SQL 등 상용 DBMS에서는 내부 데이터베이스 구조와 XML Schema와의 변환이 매우 용이한 것으로 알려져 있다. 본 연구에서는 해양지리정보 데이터베이스가 XML Schema 구조로 변환될 수 있음을 가정하고, 그림 1과 같이 분산된 해양지리정보 데이터베이스 A와 B를 XML Schema로 변환 후 스키마를 매칭, 유사도 검출, 스키마 결합으로 구성되는 스키마 통합 방법을 설계하였다.

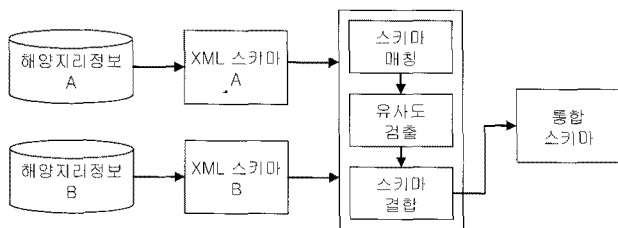


그림 1. 해양지리정보 통합 개념

Fig. 1. Concept of marine geographic data integration

본 연구에서는 위 전체 개념 중 XML Schema의 매칭 및 유사도 검출에 대해 연구하였으며, 그 사례로서 구조가 상이한 두 개의 해양GIS 메타데이터의 스키마에 적용하였다.

### 3. XML 응용스키마 통합기법

#### 3.1 유사도 측정기법

XML Schema 간의 범주와 속성을 비교하여 유사한 부분을 파악하기 위해 유사도 측정이 이루어지며, 식 (1)과 같이 교집합과 여집합의 집합 이론을 이용하여 계산된다.

$$s(a,b) = \frac{|A \cap B|}{|A \cap B| + \alpha(a,b)|A/B| + (1-\alpha(a,b))|B/A|} \quad (1)$$

for  $0 < \alpha < 1$

여기서  $a$ 와  $b$ 는 비교하고자 하는 XML Schema 항목이며,  $A$ 와  $B$ 는 항목  $a$ 와  $b$ 의 속성이다.  $\alpha$ 는 속성 계산 시 적용하는 가중치이며,  $A \cap B$ 는 공통 속성의 개수이며,  $A/B$

는  $A$ 에는 포함되나  $B$ 에는 포함되지 않는 속성의 개수,  $B/A$ 는  $B$ 에는 포함되나  $A$ 에는 포함되지 않는 속성의 개수이다. 항목  $a$ 와  $b$ 의 유사도는 비교하는 항목의 이름, 자식 항목, 부모 항목, 포함하는 속성의 종류, 형제 항목에 대해 식(1)을 이용하여 계산 후 합산하여 유사도를 계산하게 되어 식(2)와 같이 정리된다.

$$s(a,b) = \sum_{k=1}^n \delta_k s_k(a,b) \quad \sum_{k=1}^n \delta_k = 1 \quad (2)$$

먼저, 항목  $a$ 와 항목  $b$ 의 이름에 대한 유사도 측정은 식 (1)을 적용한 다음의 식으로 계산된다.

$$s_{name}(a,b) = \frac{N_{n(a) \cap n(b)}}{N_{n(a) \cap n(b)} + \alpha(a,b)N_{n(a)/n(b)} + (1-\alpha(a,b))N_{n(b)/n(a)}}$$

$$N_{n(a) \cap n(b)} = |name(a) \cap name(b)|$$

$$N_{n(a)/n(b)} = |name(a) / name(b)|$$

$$N_{n(b)/n(a)} = |name(b) / name(a)|$$

다음으로 자식 항목의 유사도 역시 식(1)을 적용한 다음 식으로 계산된다.

$$s_{children}(a,b) = \frac{N_{c(a) \cap c(b)}}{N_{c(a) \cap c(b)} + \alpha(a,b)N_{c(a)/c(b)} + (1-\alpha(a,b))N_{c(b)/c(a)}}$$

항목  $a$ ,  $b$  간의 속성과 형제 클래스의 유사도 측정 방법은 위 자식 항목 측정 방법과 동일하며, 부모 항목 측정 방법은 다음과 같이 정의된다.

$$s_{parent}(a,b) = \begin{cases} 1 & parent(a) = parent(b) \\ 0 & parent(a) \neq parent(b) \end{cases}$$

결과로, 유사도 측정값  $s(a,b)$ 은 항목명, 자식항목, 부모 항목, 속성종류, 형제항목에 대해 각각 계산하고 각 계산 값에 가중치를 적용한 합으로 계산된다.

$$s(a,b) = \delta_1 s_{name}(a,b) + \delta_2 s_{children}(a,b) + \delta_3 s_{parent}(a,b) + \delta_4 s_{attribute}(a,b) + \delta_5 s_{brother}(a,b) \quad (3)$$

여기서, 가중치  $\delta_i (i=1, \dots, 5)$ 는  $i$ 번째 가중치로서 모든 가중치의 합은 1이 된다.

$$\sum_{i=1}^5 \delta_i = 1 \quad (4)$$

비교하고자 하는 두 스키마  $C_1$ 과  $C_2$ 가 있고, 각 스키마에 포함된 항목이  $m$ 과  $n$ 개 있을 때, 유사도 측정 행렬은 다음과 같이 표현된다.

$$S = \begin{bmatrix} s(1,1) & \dots & s(1,n) \\ \vdots & \ddots & \vdots \\ s(m,1) & \dots & s(m,n) \end{bmatrix} \quad (5)$$

여기서  $s(i,j)$ 는 스키마  $C_1$ 의  $i$ 번째 항목과 스키마  $C_2$ 의  $j$ 번째 항목간의 유사도를 말한다.

#### 3.2 의미관계 정의 방법

XML Schema에는 클래스 항목의 성격을 정의할 때 복합 타입 정의 요소로서 확장을 할 수 있으며, 확장을 의미하는  $\langle xs:extension \rangle$ 과 제약을 의미하는  $\langle xs:restriction \rangle$

등이 사용된다. 확장 방법은 5가지로 구분할 수 있으며 5가지 방법은 세 가지 의미적 모델링 방법으로 압축할 수 있다. 다음은 XML Schema의 5가지 확장 방법과 의미적 관계에 대해 설명하였다.

- <xs:extension> : 신규로 정의된 복합 타입은 원래의 정보와 함께 신규 타입에 적합한 부가 정보를 포함하고 있음.
- <xs:restriction> : 신규 타입에 표현된 값은 원 정보에서 부여된 값에서 제한된 값으로 정의됨
- <xs:sequence> : 신규 타입이 정의되었을 경우 동일한 순서 제약을 가짐
- <xs:all> : 위 sequence와 달리 개별의 제약 순서를 가짐
- <xs:choice> : 배타적인 관계를 가질 경우 본 타입을 사용함

위 5가지 XML Schema의 표현 방식은 다음 표와 같이 의미적 연계관계로 정의할 수 있다.

표 1. 의미적 연결 관계 종류  
Table. 1 Semantic relationships of constructors

Types of constructors	의미적 관계		
	일반화	집합	연계
$g : \langle \text{xs:extension} \rangle$	✓		
$f : \langle \text{xs:restriction} \rangle$	✓		
$s : \langle \text{xs:sequence} \rangle$		✓	
$l : \langle \text{xs:all} \rangle$		✓	
$c : \langle \text{xs:choice} \rangle$			✓

relaxation labeling 기법은 XML Schema의 구조적 제약과 의미적 연계 관계를 이용한다. 비교하고자 하는 두 XML Schema의 의미적 관계는 다음과 같이 정의할 수 있다.

- $edge: e(v_1, v_2)$  항목 간에 직접적으로 연결되었을 경우  $e(v_1, v_2)$ 으로 정의함. 여기서  $v_1$ 은  $v_2$ 의 부모 항목이고,  $v_2$ 는  $v_1$ 의 자식 항목으로, 직접적으로 연결된 항목이 있다면,  $e(v_1, v_2)=1$ 이고 그렇지 않으면  $e(v_1, v_2)=0$
- $path: e \sim (v_1, v_n)$  항목 간에 연계 관계인  $e \sim$ 의 성립 여부를  $e \sim (v_1, v_n)$ 로 정의하며,  $v_1$ 과  $v_n$ 간의 연계 관계가 성립할 경우  $e \sim (v_1, v_n)=1$ , 그렇지 않을 경우  $e \sim (v_1, v_n)=0$ 으로 정의함. 이 때,  $e \sim (v_1, v_n)$ 의 깊이는  $d_{ij}$ 로 표기
- $edge\ type: r(v_i, v_j)$  항목  $v_i, v_j$  간에 직접적인 연결이 있을 경우  $v_i$ 와  $v_j$  간의 타입은  $\kappa(\kappa \in \{g, f, s, l, c\})$ 로 표기되며,  $r(v_i, v_j) = \kappa$ 로 정의할 수 있다. 만약  $v_i$ 와  $v_j$ 간의 직접적인 연결이 없을 경우에는  $r(v_i, v_j) = null$ 로 정의
- $path\ type: r \sim (v_1, v_n)$  항목  $v_i$ 과  $v_n$ 이 주어질 때,  $v_1$ 에서부터  $v_n$  연계 관계가 존재할 경우, 연계 관계는  $e_1, e_2, \dots, e_{n-1}$ 로 표기하며, 연계 타입은  $r_1, r_2, \dots, r_{n-1}$ 로 표기할 수 있다. 만약  $r_1 = r_2 = \dots = r_{n-1} = \kappa$ 라고 할 때  $r \sim (v_1, v_n) = \kappa$ 로 정의
- $type\ similarity: sim(r_1, r_2)$  연계관계 타입  $r_1, r_2$ 간의 의미적 유사관계는  $sim(r_1, r_2)$ 로 정의되며, 만약 두 연계관

계 타입  $r_1$ 과  $r_2$ 이 의미적 연계관계가 있으면  $sim(r_1, r_2) = 1, sim(r_1, \sim r_2) = 1, sim(r_1 \sim, r_2 \sim) = 1$ 로 정의

### 3.3 XML 스키마 의미적 연계관계 파악기법

Relaxation labeling 기법은 제약을 가진 그래프 노드에 라벨을 할당하는 문제에 적용되어 왔으나, 본 연구에서는 XML 스키마의 의미적 연계관계 파악을 위해 적용하였다. 비교하고자 하는 두 스키마를  $B = \{b_1, \dots, b_m\}$ ,  $A = \{\lambda_1, \dots, \lambda_n\}$ 라고 표현할 때, 항목 간 의미적 제약관계는 적합계수  $R_{ij} = \{r_{ij}(\lambda, \mu)\}$ 로 표현되며,  $r_{ij}(\lambda, \mu)$ 는 오브젝트  $i$ 에 라벨  $\lambda$ 가 할당되고, 오브젝트  $j$ 에 라벨  $\mu$ 가 할당될 때의 적합도를 말한다. Relaxation labeling 기법은 초기값으로 비 제약값을 가진 적합도 값인  $p_i^0(\lambda)$  값을 설정하고, 반복적으로 값을 보완하기 위해 제약 모델에 따라 지원 함수를 적용한다. 비교 하고자 하는 스키마  $C_1$ 의 항목 수가  $m$ 개, 스키마  $C_2$ 의 항목 수가  $n$ 개 있을 때, 적합계수 행렬은 다음과 같이 표현된다.

$$R = \begin{bmatrix} R_{11} & \dots & R_{1m} \\ \vdots & \ddots & \vdots \\ R_{m1} & \dots & R_{mm} \end{bmatrix} \quad (6)$$

$$R_{ij} = \begin{bmatrix} r_{ij}(1,1) & \dots & r_{ij}(1,n) \\ \vdots & \ddots & \vdots \\ r_{ij}(n,1) & \dots & r_{ij}(n,n) \end{bmatrix} \quad (7)$$

클래스  $C_1$ 의 항목  $i$ 와 클래스  $C_2$ 의 항목  $\lambda$ 의 적합도  $p_i(\lambda)$ 를 계산하기 위해 다음의 지원 함수가 요구된다.

$$q_i(\lambda) = \sum_{j=1}^m \sum_{\mu=1}^n r_{ij}(\lambda, \mu) p_j(\mu) \quad (8)$$

위 식에 따라 업데이트 식은 다음과 같이 표현된다.

$$p_i^{(t+1)}(\lambda) = f(p_i^{(t)}(\lambda), q_i^{(t)}(\lambda)) \quad (9)$$

여기서  $f$ 는 지원 함수이며,  $t$ 는 반복단계를 나타낸다. 본 연구에서는 참고문헌에 따라 지원함수  $f$ 로 다음 식을 적용하였다.

$$p_i^{(t+1)}(\lambda) = \frac{p_i^{(t)}(\lambda) q_i^{(t)}(\lambda)}{\sum_{\mu=1}^m p_i^{(t)}(\mu) q_i^{(t)}(\mu)} \quad (10)$$

클래스 항목 매칭의 적합도 행렬은 다음과 같이 정리된다.

$$P^{(t)} = \begin{bmatrix} P_1^{(t)}(1) & \dots & P_1^{(t)}(n) \\ \vdots & \ddots & \vdots \\ P_m^{(t)}(1) & \dots & P_m^{(t)}(n) \end{bmatrix} \quad (11)$$

relaxation labeling 기법에서 적합 계수는 XML Schema 문법에 따른 의미관계 정의방법에 따라 기술할 수 있는데, 항목 간 직접 연결관계는  $e(v_1, v_2)$ , 계층적인 연계

관계는  $e \sim (v_1, v_n)$ , 연결관계 타입은  $r(v_i, v_j)$ , 연계관계 타입은  $r \sim (v_1, v_n)$ , 연계관계타입유사도는  $sim(r_1, r_2)$ 으로 정의된다.

스키마  $C_1$ 의 두 항목  $v_i, v_j$ , 스키마  $C_2$ 의 두 항목  $v_\lambda, v_\mu$ 가 있을 때,  $v_i$ 가  $v_\lambda$ 에 일치하고  $v_j$ 가  $v_\mu$ 에 일치하는 적합계수 정의는 다음 식과 같다. 아래 식에서와 같이 적합계수  $r_{ij}(\lambda, \mu)$ 는 항목 간의 연결 형태인 연결  $e(v_1, v_2)$ , 연계  $e \sim (v_1, v_n)$ , 연결 형태  $r(v_i, v_j)$ , 연계 형태  $r \sim (v_1, v_n)$ 에 따라  $w_i (i=1,2,3)$ 의 가중치를 부여 받는다. 여기서  $\wedge$ 은 집합이론에서 교집합의 의미이다.

$$w_1 : \text{if } e(i, j) = 1 \wedge e(\lambda, \mu) = 1 \wedge r(i, j) = r(\lambda, \mu)$$

$$w_2 : \text{if } e(i, j) = 1 \wedge e(\lambda, \mu) = 1 \wedge \sim (r(i, j), r(\lambda, \mu)) = 1$$

$$\frac{w_1}{d_{ij} + d_{\lambda\mu}} : \text{if } e(i, j) = 1 \wedge e \sim (\lambda, \mu) = 1 \wedge r(i, j) = r \sim (\lambda, \mu) \text{ or } e \sim (i, j) = 1 \wedge e(\lambda, \mu) = 1 \wedge r \sim (i, j) = r(\lambda, \mu) \text{ or } e \sim (i, j) = 1 \wedge e \sim (\lambda, \mu) = 1 \wedge r \sim (i, j) = r \sim (\lambda, \mu)$$

$$\frac{w_2}{d_{ij} + d_{\lambda\mu}} : \text{if } e(i, j) = 1 \wedge e \sim (\lambda, \mu) = 1 \wedge \sim (r(i, j), r \sim (\lambda, \mu)) = 1 \text{ or } e \sim (i, j) = 1 \wedge e(\lambda, \mu) = 1 \wedge \sim (r \sim (i, j), r(\lambda, \mu)) = 1 \text{ or } e \sim (i, j) = 1 \wedge e \sim (\lambda, \mu) = 1 \wedge \sim (r \sim (i, j), r \sim (\lambda, \mu)) = 1$$

$$w_3 : \text{others}$$

스키마  $C_1$ 의 항목  $v_i$ 와  $v_j$ 간에 연결 관계( $e(i, j)=1$ )와 스키마  $C_2$ 의 항목  $v_\lambda$ 와  $v_\mu$ 간의 연결 관계( $e(\lambda, \mu)=1$ )가 성립하고 연결 형태인  $r(i, j)$ ,  $r(\lambda, \mu)$ 가 같을 경우 적합계수  $r_{ij}(\lambda, \mu)$ 는  $w_1$ 이 된다. 만약 연결 형태인  $r(i, j)$ ,  $r(\lambda, \mu)$ 가 다르나, 의미적 연계 형태인  $sim(r(i, j), r(\lambda, \mu)) = 1$ 일 경우에 적합계수는  $w_2$ 가 된다.  $w_2$ 는  $w_1$  보다 작으며, 본 연구에서는  $w_1=1.0$ ,  $w_2=0.8$ 을 적용하였다.

스키마  $C_1$ 에서 항목  $v_i$ ,  $v_j$ 간의 연결/연계 관계 ( $e(i, j) = 1, e \sim (i, j) = 1$ )와 스키마  $C_2$ 에서 항목  $v_\lambda$ ,  $v_\mu$ 간의 연결/연계 관계( $e(\lambda, \mu) = 1, e \sim (\lambda, \mu) = 1$ )가 성립하고, 각 항목 간의 연결, 연계 형태( $r(i, j) = r \sim (\lambda, \mu)$  or  $r \sim (i, j) = r(\lambda, \mu)$  or  $r \sim (i, j) = r \sim (\lambda, \mu)$ )가 같을 경우 적합계수는  $w_1 / (d_{ij} + d_{\lambda\mu})$ 로 정의되며, 이 때  $d_{ij}$ 와  $d_{\lambda\mu}$ 는 연계 관계의 깊이이다. 만약 연결, 연계 형태가 다르나 의미적 연계 관계( $sim(r(i, j), r \sim (\lambda, \mu)) = 1$  or  $sim(\sim(r(i, j), r(\lambda, \mu)) = 1$  or  $sim(r \sim (i, j), r \sim (\lambda, \mu)) = 1$ )가 성립되면 적합계수는  $w_2 / (d_{ij} + d_{\lambda\mu})$ 로 정의된다. 만약 스키마  $C_1$ 의 항목  $v_i$ 와  $v_j$ 간의 연결/연계 관계가 없으며, 스키마  $C_2$ 의 항목  $v_\lambda$ 와  $v_\mu$ 간의 연결/연계 관계가 없을 경우에는 적합계수는  $w_3$ 으로 정의된다.

## 4. 해양GIS 메타데이터의 통합

### 4.1 해양GIS 메타데이터

메타데이터(Metadata)는 “데이터에 대한 데이터” 즉, 데이터에 대한 이력서로서 지리정보에 대한 이해를 높이고 정보의 활용을 촉진하는 중요한 기능을 담당하고 있다. 현재 GIS 및 RS 데이터베이스 구축에 소요되는 엄청난 비용을 감안하면 한번 수집된 지리정보의 공동 활용에 대한 중요성은 매우 크다. 지리정보 메타데이터는 지형공간정보에 대한 내용, 품질, 용도 등 상세한 정보를 사전에 제공함으로써 사용자의 요구에 맞는 정보의 접근을 용이하게 하고 불필요한 지리정보의 송수신 과정을 간소화하여 지리정보 유통의 효율성을 제고시킬 수 있다. 해양GIS 메타데이터는 다음의 기

능이 있다.

- 데이터 검색 : 온라인상의 검색 및 관람을 위한 내용, 품질, 연락정보, 공급처, URL 등에 대한 요약 정보
- 데이터 사용 : 데이터 적용범위, 유지관리, 내용, 데이터 생성에 대한 보다 넓은 범위에 대한 정보
- 데이터 적합도 : 사용, 제한, 포맷, 연대, 범위에 관한 부가적인 정보. 본 수준의 메타데이터 정보는 사용자에게 데이터 적합도를 검증할 수 있는 정보를 제공한다.
- 데이터 공유 : 데이터 내용, 교환 포맷, 공간 표현에 관한 보다 자세한 정보
- 데이터 유지관리 : 데이터 품질절차 및 테스트 결과 등을 포함한 가장 상세한 수준의 정보. 본 수준의 정보는 조직 간의 데이터 교환이 이루어질 때 중요하다.

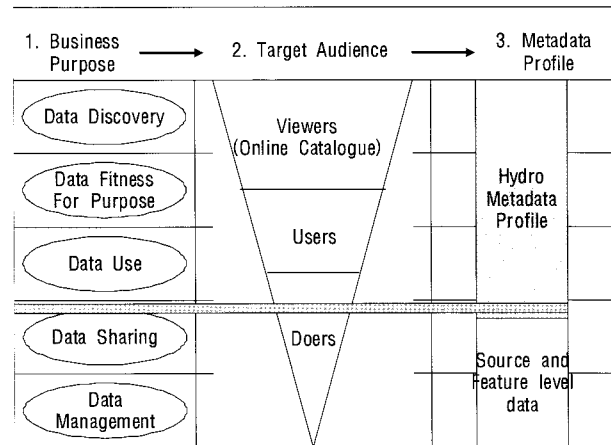


그림 2. 메타데이터 기능  
Fig. 2. Functions of Metadata

국제표준기구에서는 해양GIS 메타데이터의 데이터 검색, 사용, 적합도 검증, 공유, 유지관리 등 다양한 사용 목적을 위해 수백 개의 항목과 복잡한 구조로 구성된 메타데이터 스키마를 표준으로 제공하여, 각 국의 관련 기관에서 표준의 프로파일링 기법을 통해 표준을 사용하도록 하였다. 그러나 메타데이터 스키마 항목의 복잡한 구조로 인해 표준에서 정한 프로파일링 기법 적용 없이 개별의 메타데이터 구조를 채택하여 사용하여 기관 간의 메타데이터 공유가 이루어지지 못하였다.

### 4.2 해양GIS 메타데이터의 통합

본 연구에서는 스키마 통합 기법 적용을 위해 해양GIS 메타데이터 사례를 선정하여 적용하였다. 그림 3은 연락처에 관한 스키마와 데이터 식별에 관한 스키마로서 해양지리정보의 생산자의 연락처와 관리 도업에 관한 정보를 포함하고 있다. 그림 3의 연락처 스키마는 담당자, 직위, 주소, 전화, 관리도업의 항목으로 구성되며, 데이터식별 스키마에는 검색정보, 연락처, 유통정보, 개인명, 기관명, 도업정보, 지역범위, 연락정보, 도업번호, 행정경계, 주소, 온라인, 전화의 항목으로 구성되어 있다.

그림 3의 스키마의 개별 항목의 유사도를 측정하기 위해 식 (1)을 이용하여 다음과 같이 항목명, 자식항목, 부모항목, 속성종류, 형제항목을 계산하였다.

$$s(a, b) = \delta_1 s_{name}(a, b) + \delta_2 s_{children}(a, b) + \delta_3 s_{parent}(a, b) + \delta_4 s_{aribute}(a, b) + \delta_5 s_{brother}(a, b)$$

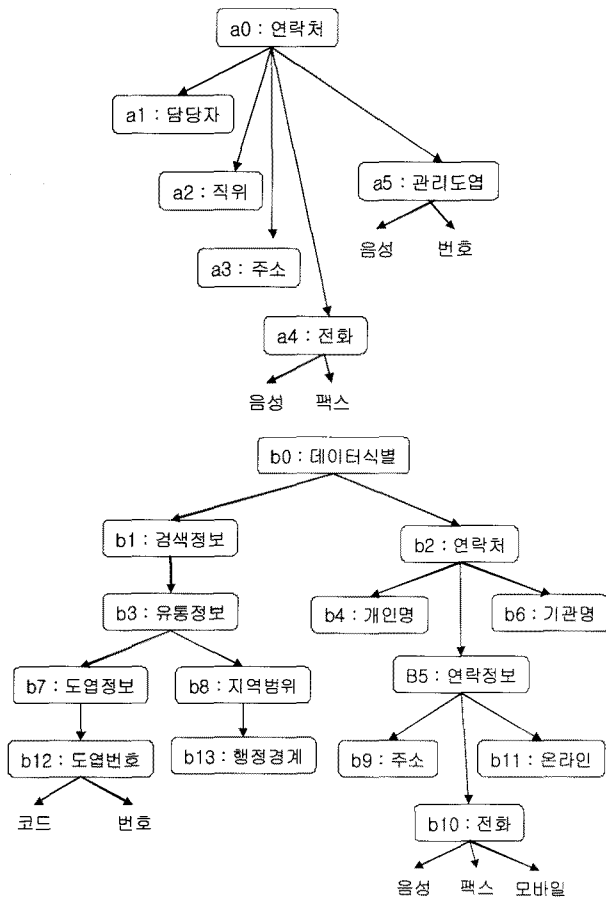


그림 3. 메타데이터 스키마 구조  
Fig. 3. Structure of Metadata Schema

본 연구에서는 각 계산의 가중치는  $\delta_1=0.2, \delta_2=0.3, \delta_3=0.2, \delta_4=0.2, \delta_5=0.1$ 으로 적용하여 초기 비제약식 유사도 측정값인  $p_i^0(\lambda)$  행렬을 계산하였으며, 식(6)~식(11)에 따라 의미적 연계관계를 고려한 적합도 행렬을 계산하였다. 본 연구의 계산 절차는 다음과 같다.

```

input  $s(i, \lambda); t = 0;$ 
being  $p_i^{(0)}(\lambda) = s(i, \lambda);$ 
while (not stop_condition) do
    evaluate  $q_i^{(t)}(\lambda) = \sum_{j=1}^m \sum_{\mu=1}^n r_{ij}(\lambda, \mu) p_j^{(t)}(\mu);$ 
     $p_i^{(t+1)}(\lambda) = f(p_i^{(t)}(\lambda), q_i^{(t)}(\lambda));$ 
     $t = t + 1;$ 
end while;
output  $p_i^{(t)}(\lambda);$ 
end
    
```

본 알고리즘의 종료 조건은 단계  $t$ 와 단계  $t+1$  간의 적합도  $p_i^{(t)}(\lambda)$  차이에 의해 결정된다. 최종 단계와 이전 단계의 차이인  $|p_i^{(t+1)}(\lambda) - p_i^{(t)}(\lambda)|$ 가 한계치보다 적을 때 종료된다.  $p_i^{(t)}(\lambda)$ 은  $t$  단계에서 스키마  $C_2$ 의 항목  $v_j$ 에 대응되는 스키마  $C_1$ 의 항목  $v_i$ 의 적합도 값을 나타낸다. 위 식의 계산결과로 클래스 항목의 적합도 뿐만 아니라 의미적 관계를 고려한 유사도를 산출하였으며, 두 스키마 간의 최적 매칭 결과를 도출하였다.

알고리즘에 따른 계산 결과 3단계인  $t=3$ 에서 유효한 값을 도출하였으며, 각 항목에서 의미적 유사관계가 적용된 유사도 측정값 중 최고값을 선별한 결과  $a_0 \leftrightarrow b_5$ (연락처 $\leftrightarrow$ 연락정보),  $a_3 \leftrightarrow b_9$ (주소 $\leftrightarrow$ 주소),  $a_4 \leftrightarrow b_{10}$ (전화 $\leftrightarrow$ 전화),  $a_5 \leftrightarrow b_{12}$ (관리도입 $\leftrightarrow$ 도입번호) 항목에서 유사도 값이 높게 계산되었다. 특히, 초기 비제약 유사도 측정 시  $a_0$ 인 관리도입이  $b_7$ 인 도입정보와 유사한 것으로 계산되었으나 relaxation la-

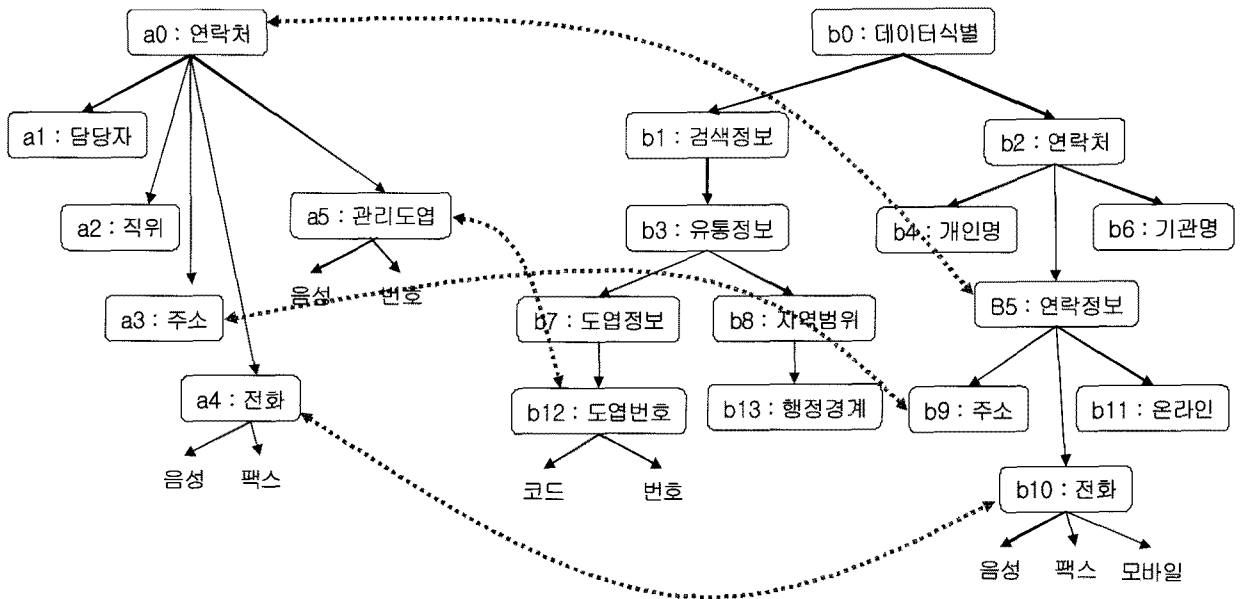


그림 4. 스키마 항목 비교 및 유사항목 도출  
Fig. 4. Comparison of Schema elements and Calculation of similarity elements

beling 기법을 이용한 의미적 유사관계 가중치를 적용하여  $b_5$ 의 유사도가 높게 계산되는 것을 알 수 있었다.

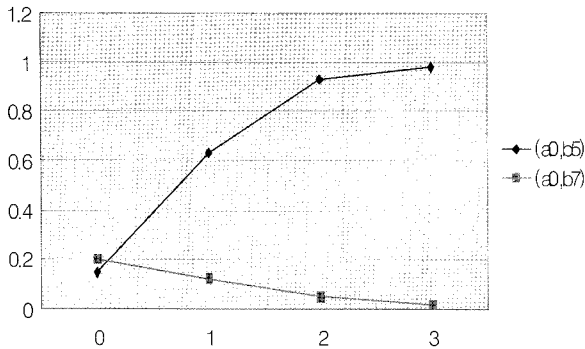


그림 5. 유사도 측정값의 변화  
Fig. 5. Changes of Similarity measure

### 5. 결 론

본 연구에서는 해양지리정보 DB 통합을 위해 XML 응용스키마 간의 항목 매칭을 수행하였다. XML 응용스키마의 의미 모델은 클래스 항목, 속성, 연관관계 및 기록, 결합관계 등의 구조적, 의미적 내용을 표현한다. 비교하고자 하는 두 스키마의 항목 간 유사도 측정기법은 클래스 항목명, 속성, 이웃 항목으로 계산되며, 항목 간의 의미적 결합관계를 고려하기 위해 relaxation labeling 기법이 적용되었다.

해양지리정보 DB 통합을 위해서는 관계형 DB의 XML 응용스키마 변환, 스키마 간의 매칭 및 유사도 측정, 스키마 결합 및 통합 스키마 생성 과정이 필요하며, 본 연구에서는 위 절차 중 스키마를 매칭하여 유사도를 측정하였다. 앞으로 클래스 항목 간 비교 시 온톨로지 개념 적용과 유사한 클래스 간 스키마를 자동결합 할 수 있는 방법 개발이 요구된다.

### 참 고 문 헌

- [1] Marcello Pellillo(1994), "Learning Comaptibility Coefficients for Relaxation Labeling Processes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 9. pp.933-945.
- [2] Qin Chen, J.Y.S.Luh(1995), "Relaxation labeling Algorithm for information integration and its convergence", *Pattern Recognition*, Vol. 28, No. 11. pp.1705-1722.
- [3] Marcello Pellillo(1997), "The Dynamics of Nonlinear Relaxation Labeling Processes", *Journal of Mathematical Imaaing and Vision*, Vol. 7, pp. 309-323.
- [4] Andrea torsello, Edwon R. Hancock(2003), "Computing approximate tree edit distance using relaxation labeling", *Pattern Recognition Letters* 24, pp. 1089-1097.

### 저 자 소개



오세웅(Se-Woong Oh)  
1999년 : 한국해양대학교 물류시스템공학과 졸업  
2001년 : 동 대학원 석사 졸업  
2008년 : 동 대학원 박사 졸업

관심분야 : 해양GIS, 전자해도, 데이터베이스 통합, 퍼지이론  
Phone : 042-868-7297  
Fax : 042-869-0435  
E-mail : osw@moeri.re.kr



박계각(Gyei-Kark Park)  
1982년 : 한국해양대학교 항해학과 졸업  
1989년 : 동 대학원 석사 졸업  
1993년 : 동경공업대학교 대학원 박사 졸업

관심분야 : 지능시스템, 해상정보시스템, 항만 물류 및 국제 경제학  
Phone : 061-240-7164  
Fax : 061-240-2344  
E-mail : gkpark@mmu.ac.kr



박종민(Jong-Min Park)  
1994년 : 경북대학교 전자계산학과 졸업  
1996년 : 동 대학원 석사 졸업

관심분야 : 해양GIS, 인공지능, 다이나믹 모델링  
Phone : 042-868-7259  
Fax : 042 869-0435  
E-mail : pjmm@moeri.re.kr



서상현(Sang-Hyun Suh)  
1979년 : 서울대학교 조선공학과 졸업  
1982년 : 동 대학원 석사 졸업  
1991년 : 미시간 대학교 박사 졸업

관심분야 : 해양GIS, 전자해도, 조선공학, 퍼지이론, GNSS  
Phone : 042-868-7264  
Fax : 042-868-7503  
E-mail : shsuh@moeri.re.kr