2005

Bridging the Gap between Flexibility and Rigidity: Unifying Relationships with a Hybrid Link Architecture

Yiwen Zhang
University of Arizona, yiwen@u.arizona.edu

Mohan Tanniru
University of Arizona, mtannriu@eller.arizona.edu

Follow this and additional works at: http://aisel.aisnet.org/amcis2005

Recommended Citation
http://aisel.aisnet.org/amcis2005/345

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2005 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
**Bridging the Gap between Flexibility and Rigidity: Unifying Relationships with a Hybrid Link Architecture**

Yiwen Zhang  
Department of Management Information Systems  
University of Arizona  
yiwen@u.arizona.edu

Mohan Tanniru  
Department of Management Information Systems  
University of Arizona  
mTanniru@eller.arizona.edu

**ABSTRACT**

Current knowledge systems lack an effective architecture to bridge the well-acknowledged gap between flexibility and rigidity - users' preference of flexible expression and machine's need for rigorous representation. This paper focuses on finding an adequate set of relationship types (or link types) and map flexible expressions from users to link types in propositional knowledge systems, in which knowledge elements are stored as “concept-relationship-concept” triplets. We call our approach “hybrid link architecture”. This architecture has two levels of presentations, an open link name layer and a closed link type layer. Between these two layers is a matching mechanism based on a psycholinguistic thesaurus to map names to types. The link types are derived by synthesizing literature of knowledge organization, semantic network, and educational taxonomies. A two-stage evaluation is conducted based on 39,706 triplets. The evaluation shows that the matching mechanism is accurate and the proposed link types are mutually exclusive and cover most link names in our dataset.

**KEYWORDS**: propositional knowledge system, relationship name/link name, relationship type/link type

**1. INTRODUCTION**

Researchers summarize current challenges in knowledge management strategy and practice in a recent review (Alavi and Leidner 2001). Those challenges include how to capture the knowledge from individuals, how to make individual knowledge meaningful to others, and how to integrate the knowledge from individuals to generate organizational level wisdom. They call for both a strategy and a set of processes that can capture, present, and codify knowledge from individuals, so that the knowledge can be shared, integrated, and reused.

A promising solution is to utilize propositional knowledge system, which uses “concept – relationship–concept” or “node-link-node” triplets as basic knowledge elements. These systems, especially the ones using graphical interface, have gained popularity in the current knowledge management practice. For example, concept maps, also called cognitive maps, mind maps, and knowledge maps, have been increasingly deployed in knowledge acquisition, representation, organization, and sharing and reuse (Gaines and Shaw 1995; Aidman and Egan 1998; Novak 1998; Leake, Maguitman et al. 2002; Merrill 2002).

However, most propositional knowledge systems do not have an effective design to address the differences between users’ preference of flexible expression and machine’s need for rigorous representation. Users prefer flexible expressions in presenting concepts and relationships; while machines can only operate on inputs in rigorous form to perform knowledge inference and integration. This paper focuses on the expression and presentation of “relationships”. Given that relationships are often presented by and referred to as “links” in computer systems, we use “link” and “relationship” inter-changeably from now on. Current systems use either ‘open’ or ‘closed’ link architecture. The “open” link architecture allows users to put in any phrase, and the ‘closed’ link architecture requires users to select from a predefined set of phrases. Both architectures cannot address the gap between flexibility and rigidity, which motivates us to propose a hybrid link architecture.

The rest of the paper is organized as follows. In Section 2, we discuss the limitations of closed and open link architectures, and present the design of a hybrid architecture, which has an open link name layer, a closed link type layer, and a matching mechanism to map names to types. In Section 3, we derive a closed set of link types by synthesizing literature in knowledge organization, semantic network, and educational taxonomies. In Section 4, we describe the matching process by working through an example. In Section 5, a two-stage evaluation is conducted based on 39,706 triplets generated by more than 100 graduate students during their course work. We conclude the paper in Section 6 with discussion and future work.

**2. DESIGNED A HYBRID LINK ARCHITECTURE**
2.1. Limitations of the current link architectures

Open link architecture is flexible to users, but machines can not detect the similarities and differences among potentially unlimited expressions. In addition, open link architecture is shown to be of little help to knowledge acquisition process, as it does not provide potential relationships that can guide knowledge exploration (Kinchin 2000). Open link architecture may also lead to misunderstanding during knowledge sharing.

Closed link architecture could be rigorous with limited and predefined phrases, but users may not find appropriate expressions from the closed set. These architectures are also too system-specific in that different systems have different sets of phrases. This makes it difficult to integrate knowledge from multiple systems (Guarino 1998).

2.2. A hybrid link architecture

We propose a hybrid link architecture, which combines the strengths and overcomes the weakness of both open and closed link architectures. The proposed architecture has a two-level and semi-open structure. Each link has a “link name”, referring to the phrase itself, and a “link type”, referring to an abstraction of the semantic relationship the phrase represents. The “link type” is closed and includes a predefined set of link types. The “link name” is open, and users have the freedom to use the phrases they are comfortable with and conform to the naming conventions of the specific domain if there is any.

A link name provided by users has to be mapped to a link type before it is processed by a machine for integration and reuse. The mapping process requires a systematic matching mechanism. Figure 1 illustrates the basic structure of the hybrid link architecture. The matching mechanism in the architecture should have explicit and programmable rules, so that it can be implemented easily and do the matching consistently. The set of link types and the matching mechanism are presented in the next two sections.

Figure 1. Structure of the Hybrid Link Architecture

3. DEFINING LINK TYPES

Although there is an increasing interest in identifying and utilizing relationships for knowledge management and engineering, especially in the recent research on ontology and information systems research (Guarino 1998; Leroy and Chen 2001; Merrill 2002), no broadly acceptable link types have been defined. We identified a set of semantically meaningful link types from three streams of research: knowledge organization, semantic network, and educational taxonomy. The results are presented in Figure 2 and explained as follows.

Figure 2. Link Types for the Hybrid Link Architecture
3.1. Three most fundamental link types in knowledge organization

From the literature in knowledge organization, we identified “hierarchical” (or “is-a”), “componential” (or “part-of”), and “causal” (or “cause-effect”) relationships as three most fundamental link types. Hierarchical and componential relationships are considered by researchers, such as (Lyons 1968), the most widely used relationships to understand the world. Hierarchical relationship connects subclass concepts to super-class concepts to build taxonomies. Componential relationship decomposes an integrated physical or conceptual object into parts. Note that both hierarchical and componential relationship can be further categorized into two sub-categories based on the directions of relationship. For example, for componential relationship, we can have link from whole to part or from part to whole.

Besides these two relationships, causal relationship is also widely used. It is considered the most important relationship in decision-making and problem solving (Keil 1989). One evidence of this is that our language uses a large number of causative nouns (such as “reason”) verbs (such as “lead to”), and adverbs (such as “therefore”).

3.2. Supplementary link types from semantic network

From the literature in semantic network, we identified three additional link types, “comparative”, “sequential”, and “approximal” from Collin and Quillian’s work (Collins and Quillian 1969; Collins and Quillian 1972). Semantic network research studies the structure and storage of human memory, and Collin and Quillian’s work is considered very significant in the literature.

During the 60’s and 70’s, semantic network research had generated a large number of various link types. For example, Collin and Quillian’s had defined more than ten link types. Our selection of link types was guided by Brachman’s seminal work (Brachman 1979) on the classification of link types in semantic network research. Brachman pointed out that researchers defined various link types for various purposes. For example, some link types were designed only for logical or implementational purpose, such as “conjunctive” and “disjunctive”. We identified link types that were used for conceptual knowledge expression from Collin and Quillian’s work. They were “comparative”, “sequential”, and “approximal”, besides the three link types we have already identified in 3.1.

3.3. Other link types suggested by educational taxonomies

We further defined “descriptive” link type and proposed five sub-types. They were “explanation”, “application”, “characterization”, “evaluation”, and “operation”. They were mainly adapted from Bloom’s taxonomy of educational objectives (Bloom, Englehart et al. 1956).

Bloom’s taxonomy is of significant theoretical and practical importance in the education and learning literature. For example, it has been widely used to guide learning process and question design. This taxonomy suggested six levels of objectives in a student’s learning process. The first two columns in Table 1 list these six levels and their descriptions. We mapped these levels of objectives into our previously defined link types, and found that “analysis”, “synthesis”, and part of the “comprehension” can map to some of the previously identified link types. The corresponding link types for these levels are listed in the third column of Table 1 in italic font. The remaining levels turned out to be related to descriptive information of a focal concept, and we introduced “descriptive” link type. “Application”, “characterization”, “explanation”, and “evaluation” were introduced as four sub-link types.

<table>
<thead>
<tr>
<th>Level of objectives</th>
<th>Description</th>
<th>Corresponding link types and sub-link types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminology</td>
<td>Recall of facts</td>
<td>Explanation, characterization</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Understanding the meaning, translation, interpolation, and interpretation of problems</td>
<td>Comparative, Explanation</td>
</tr>
<tr>
<td>Application</td>
<td>Using a concept in a new situation or unprompted use of an abstraction</td>
<td>Application</td>
</tr>
<tr>
<td>Analysis</td>
<td>Separating material or concepts into component parts to understand its organizational structure</td>
<td>Componential (from whole to part), Hierarchical (from upper level to lower level)</td>
</tr>
<tr>
<td>Synthesis</td>
<td>Building a structure or pattern from diverse elements</td>
<td>Componential (from part to whole), Hierarchical (from lower level to upper level)</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Making judgments about the value of ideas or materials</td>
<td>Evaluation</td>
</tr>
</tbody>
</table>

Table 1. Bloom’s Taxonomy and Corresponding Link Types

There were a fairly large number of verbs that do not imply strong semantic meanings, and we introduced another sub-link type, “operation”, to capture this relationship. In summary, these five sub-types of “descriptive” link type are listed as follows.
3.4. Summary of link types

All the link types we identified above are listed in Table 2 with descriptions and examples. We also identified most frequently used link names from relevant literature of each identified link type. Furthermore, several closed link architectures were used to pre-examine the coverage of the newly defined link types, as illustrated by Table 3. They were proposed by Dansereau and Holley (Dansereau and Holley 1982), Lambiotte et al. (Lambiotte, Dansereau et al. 1989), and Harmon and Dinsmore (Harmon and Dinsmore 1994), respectively. We found that our link types cover all of the link types in those closed link architectures as demonstrated in Table 3.

<table>
<thead>
<tr>
<th>Link type/ sub link type</th>
<th>Description</th>
<th>Suggested link names</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical</td>
<td>One concept is a sub class/category or a specific example of the other concept.</td>
<td>example of, type of, member of; (in reverse: super type of)</td>
<td>dog -&gt; animal relational database -&gt; database</td>
</tr>
<tr>
<td>Componential</td>
<td>One concept was a component of the other one.</td>
<td>part of, component of; (in reverse: consist of, contain, compose)</td>
<td>leg -&gt; body operation system -&gt; computer environment</td>
</tr>
<tr>
<td>Causal</td>
<td>One concept is the result of the other concept.</td>
<td>prior to, next; (in reverse: followed by)</td>
<td>heat -&gt; melting fog -&gt; low visibility</td>
</tr>
<tr>
<td>Comparative</td>
<td>Two concepts have similar or opposite values or features in some dimensions.</td>
<td>like, similar to; (in reverse: compare to, not like)</td>
<td>virtual classroom -&gt; traditional classroom</td>
</tr>
<tr>
<td>Sequential</td>
<td>One concept occurs chronologically soon before or soon after the other.</td>
<td>change, lead to, cause, influence</td>
<td>compiling -&gt; running</td>
</tr>
<tr>
<td>Approximal</td>
<td>One concept is located closely to the other concept.</td>
<td>adjacent to, close to,</td>
<td>America -&gt; Canada</td>
</tr>
<tr>
<td>Descriptive Application</td>
<td>One concept is the application or usage of the other concept.</td>
<td>apply, use</td>
<td>data mining -&gt; prediction</td>
</tr>
<tr>
<td>Operation</td>
<td>One concept is a direct object of the action of the other concept.</td>
<td>operate</td>
<td>Concept mapping tool -&gt; concept maps</td>
</tr>
<tr>
<td>Explanation</td>
<td>One concept is a detailed explanation of the other concept or one concept uses a different aspect to state the other concept.</td>
<td>describe, discuss</td>
<td>Bloom -&gt; an educator and psychologist</td>
</tr>
<tr>
<td>Characterization</td>
<td>One concept is a property of the other concept.</td>
<td>feature, characteristics</td>
<td>apple -&gt; round</td>
</tr>
<tr>
<td>Evaluation</td>
<td>One concept is a comment or evaluation of the other concept.</td>
<td>advantage, disadvantage</td>
<td>bubble sort algorithm -&gt; efficient</td>
</tr>
</tbody>
</table>

Table 2. Link Types and Suggested Link Names in the Hybrid Link Architecture
Table 3. Matching Link Types from Three Closed Link Architectures to Our Link Types

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical</td>
<td>Hierarchical (example of)</td>
<td>Static (type)</td>
<td>Exemplary</td>
</tr>
<tr>
<td></td>
<td>Clustering (evidence of)</td>
<td>Instructional (example)</td>
<td></td>
</tr>
<tr>
<td>Componential</td>
<td>Hierarchical (part of)</td>
<td>Static (part)</td>
<td>Componential</td>
</tr>
<tr>
<td>Causal</td>
<td>Chaining *</td>
<td>Dynamic (influences, leads to)</td>
<td>Causal</td>
</tr>
<tr>
<td></td>
<td>(lead to, results in, produces)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparative</td>
<td>Clustering (is like)</td>
<td>Instructional (analogy)</td>
<td>Similar comparative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Opposite comparative</td>
</tr>
<tr>
<td>Sequential</td>
<td>The same as *</td>
<td>Dynamic (next)</td>
<td>Sequential</td>
</tr>
<tr>
<td>Approximatal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descriptive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explanation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characterization</td>
<td>Clustering (property of)</td>
<td>Static (characteristic)</td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
<td>Instructional (comment)</td>
<td></td>
</tr>
</tbody>
</table>

4. MATCHING MECHANISM

The matching mechanism is matching rules and processes that a link name is mapped to a link type. We used a special lexicon, Wordnet, in this process. Wordnet is based on psycholinguistic theories and organizes words into synonym sets (Miller 1995). Two terms within one synonym set are considered to have the same semantic meaning.

A three-step matching process is described below. To illustrate this process, we use the mapping from a link name “mainly comprised of” to the link type “componential” as an example, and illustrate it in Figure 3. One critical operation in this matching is to extract the cue phrase from a phrase. In this context, the cue phrase is the most important verb in the phrase.

**Step 1:** Building a cue phrase set for each link type.

We use link type and suggested link names as “seeds”. We first extract cue phrases from these seeds and then extend the cue phrase set by adding synonyms of the cue phrases from WordNet. Assuming that we have “componential” link type, and “consist of” is a suggested link name for that link type. We extract “consist” from “consist of” and “component” from “componential”. For “consist”, we get “comprise” and “compose” by looking up synonyms from Wordnet.

**Step 2:** Extracting cue phrases from the link names provided by users.

“Comprise” is extracted from the link name “is mainly comprised of”.

**Step 3:** Assigning link types by matching cue phrases extracted from link names with cue phrase sets of link types.

The cue phrase “comprise” extracted from “mainly comprised of” (as in step 2) is in the cue phrase set for “componential” link type (as in step 1), so the link type “componential” is assigned to the link name “mainly comprised of”.

---

Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, NE, USA August 11th-14th 2005 1787
5. EVALUATION AND DISCUSSION

Two experiments were conducted to evaluate the proposed architecture. The first experiment evaluated the accuracy of the matching mechanism, and the second experiment evaluated the coverage of the proposed link types and the mutual exclusivity among those link types.

The data were collected from the database of the GetSmart Concept Mapping Tool (GS-CMT) (for more information, please refer to (Marshall, Zhang et al. 2003)). This tool provides graphical interface and allow users to input their knowledge as a node-link network. Each “node-link-node” element is a “concept-relationship-concept” triplet. This dataset had 1,435 concept maps, created by more than 100 graduate students. Those concept maps had a total of 39,706 triplets. The topics of these concept maps covered highly-structured lecture materials as well as open topics from assigned readings. All these data were stored in a Microsoft SQL Server, and the matching process was coded into stored procedure to facilitate automation.

5.1. Accuracy of the matching mechanism

The accuracy of the matching mechanism was evaluated by comparing the link types assigned by experts and by the matching mechanism. We asked three experts to assign link types to link names. The experts were graduate students who scored high in the course from which the dataset were collected. To cover as many diverse links as possible, we selected 5 maps created by different authors and from different topics. We presented our description of link types to experts and asked the experts to mark the link types for all the links on the maps. There were 145 links on the selected 5 maps.

Accuracy was calculated using 5-1. Since multiple experts were involved, we used the Kappa statistic, K, to measure inter-rater reliability (Carletta 1996) as defined in 5-2, where P(A) is the fraction of times that the experts agree and the P(E) is the fraction of times that they are expected to agree by chance. A Kappa score greater than 0.80 indicates high reliability (Carletta 1996).

\[
\text{Accuracy} = \frac{\text{no. of links that experts and matching mechanism assigned the same link type}}{\text{no. of links}} \quad (5-1)
\]

\[
K = \frac{P(A) - P(E)}{1 - P(E)} \quad (5-2)
\]

Results and discussion

The accuracy was 0.84 and the Kappa score was 0.86. We further looked into the cases where the agreements between the experts and the matching mechanism were not reached. We categorized them into four types, as illustrated in Table 4. For each link name, the matching mechanism might successfully assign or fail to assign it a link type. The three experts might assign it the same link type, in which case the agreement among expert was reached. They might assign different link types to a link, in which case the agreement among experts was not reached. Only when the following three conditions were reached at the same time, was the case counted as accurate, 1) the matching mechanism successfully assigned one link type to the link; 2) all the experts assigned the link the same link type; 3) the link types from 1) and from 2) were the same.
Zhang & Tanniru

Unifying Relationships with a Hybrid Link Architecture

<table>
<thead>
<tr>
<th>Matching mechanism</th>
<th>Able to assigned a link type</th>
<th>Same</th>
<th>84% of cases</th>
<th>4</th>
<th>Different</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fail to assign a link type</td>
<td>Type 2</td>
<td>Type 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Link Type Assignment -- by Experts vs. by Matching Mechanism

From further investigation, we found that Type 1, 2, and 3 accounted for most of the remaining 16% of the cases. Among the cases of Type 1, we found that those link names were either too “short” to infer a meaningful link type, or too “long” to carry one link type only. For example, some links just used a single preposition, such as “of” and “with”. These single prepositions often have many meanings. Some link names were composed of multiple cue phrases, and it was difficult to locate the most important one. In these cases, even the experts couldn’t agree on a link type. This might imply that the students were not clear about the relationships between the two concepts, and for the benefit of learning, they should think it over and restructure the triplets.

Among the cases of Type 2, we found that most link names were verbs. There were no mappings for the verbs such as “search” and “index” in the cue phrase set. However, the assigned link types by experts were consistent. For example, “index” and “search” were categorized into “operation” sub-type. It was because of the limited coverage of the derived cue phrases on verbs. Since verbs are very diverse and highly domain-specific, we cannot enumerate all of them in advance. We may improve the link system by customization, such as identifying the most important verbs in a specific domain.

If the students can avoid Type 1 cases and we can put some verbs in Type 2 cases into our mapping rules, the accuracy of this matching mechanism was expected to improve. We recalculated the accuracy and Kappa score, and they improved to 0.93 and 0.90 respectively. The results are presented in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Kappa Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td>Results after improving</td>
<td>0.93</td>
<td>0.90</td>
</tr>
<tr>
<td>Type 1 &amp; 2 cases</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Accuracy and Kappa Score of the Matching Mechanism

Type 3 cases were those link names that the experts could not agree on the same link type. We found that was caused by the semantic overlapping among evaluation, explanation, and characterization, which were sub-types of descriptive link types. The distinctions among them were not as significant as the distinctions among others. Slightly different perspectives held by different experts might lead to different assignments. For example, a feature of a concept as represented by “characterization” can be an advantage of that concept, as represented by “evaluation”. One solution to avoid such overlapping is to combine these subtypes; but by doing this we may sacrifice granularity. But trade-off is that lowering the granularity may lowering the cognitive benefit of building concept maps.

5.2. Coverage and exclusivity of the link types

“Coverage” measures to what degree the link types cover all the relationships used by the subjects; and “exclusivity” measures whether each link name only matches to one link type. Since the evaluation of the matching mechanism showed high accuracy and reliability, our approach for this evaluation was to use the matching mechanism to assign link types to all the link names in the dataset, and calculated the coverage and exclusivity as defined in 5-3 and 5-4.

\[
\text{Coverage} = 1 - \frac{\text{the number of links that can't be categorized into any proposed link types}}{\text{the total number of links}} \quad (5-3)
\]

\[
\text{Exclusivity} = \frac{\text{the number of links that can be categorized into one and only one link type}}{\text{the number of links that can be categorized into one or more link types}} \quad (5-4)
\]
There were 39,706 triplets in the dataset. Two preprocessing steps, “filtering” and “aggregation”, were conducted before the automatic matching. “Filtering” removed unlabelled and “bad” links, which accounted for 4.0% of the total number. “Bad” links were those links with single prepositions or long link names, which were usually not meaningful as we described in 4.1. Of the 4.0% removed links, unlabeled links accounted for 1.4% and “bad” links accounted for the rest of it, with single prepositions at 1.8% and long link names at 0.8%. 38,118 links were left after filtering. “Aggregation” aggregated identical names and summarized them into 5,327 distinct link names and their frequencies. The matching process then extracted 347 distinct cue phrases from the 5,327 link names, and mapped them into the link types.

**Results and discussion**

Table 6 shows the results of evaluation. The coverage of our link type set was 84% according to Formula (5-3), and 16% of link names could not be mapped to a specific link types. The mutual exclusivity was 100% since the filtering process guarantees that there is exactly one cue phrase for one link name, and in no scenario does one cue phrase belong to more than one link type (no overlap between cue phrase sets of predefined link types).

<table>
<thead>
<tr>
<th>Link Types</th>
<th>Number of Links</th>
<th>Percentage (%)</th>
<th>Cue Phrase Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical</td>
<td>8309</td>
<td>21.80</td>
<td>example, such as, case, type, member, is</td>
</tr>
<tr>
<td>Componential</td>
<td>10,464</td>
<td>27.45</td>
<td>have, consist, contain, include, compose, part, made of</td>
</tr>
<tr>
<td>Causal</td>
<td>2,340</td>
<td>6.14</td>
<td>lead to, cause, influence, determine</td>
</tr>
<tr>
<td>Comparative</td>
<td>1,853</td>
<td>4.86</td>
<td>like, compare, similar, differ</td>
</tr>
<tr>
<td>Sequential</td>
<td>1,109</td>
<td>2.91</td>
<td>next, go to, procedure</td>
</tr>
<tr>
<td>Approximal</td>
<td>0</td>
<td>0</td>
<td>close to, neighbor</td>
</tr>
<tr>
<td>Descriptive</td>
<td>7,891</td>
<td>20.70</td>
<td>application, operation, characterization, evaluation, explanation</td>
</tr>
<tr>
<td>Uncategorized</td>
<td>6,152</td>
<td>16.14</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Coverage of the Link Types

We further examined the 16% uncategorized links. Most of them were verbs. We summarized new matching rules between verbs and link types from Type 3 cases in the first evaluation. The coverage was raised to 93% after adding these rules.

6. CONCLUSIONS AND FUTURE WORK

6.1. Relevant work

Ontology and information system research involves large amount work on the categorization, formalization, and utilization of relationships and concepts. Originally, most research was conducted by psychologist, computational linguistics, and knowledge engineers. From earlier 90’s, there has been increasing interest among researchers from various domains. The context of ontology research has spread to enterprise modeling, e-commerce, and emerging web ontology language standard. One of the most significant works is in bioinformatics domain. Many famous ontology has been build and widely used in this area, such as UMLS, GO, HOGO. Expert has defined 135 concept types and 54 link types in UMLS(2005).

However, most of the works focus on concepts rather than relationships, and usually take “top-down” approach, heavily relying on the expert knowledge. Our research combined both “top-down” approach, in proposing relationship types from theories and taxonomies from literature, and a “bottom-up” approach, in collecting data from nonexperts to evaluate the research.

6.2. Contribution and future work

This paper proposed, operationalized, and evaluated a new design of link architecture for propositional knowledge systems. With this architecture, we bridge the gap between users’ need for flexible expression and machine’s need for rigorous representation. Future research will implement this link architecture in real systems, and design integration rules to explore new applications. Together with the parallel research on the matching among concepts, more interesting research can be conceived. For example, we could integrate the knowledge from different users with the same topic to create a comprehensive knowledge network or discover discrepancies among users. We could also use the explicit and implicit meaning of relationships, and build relationship-enhanced searching or “Question and Answering” systems.
7. ACKNOWLEDGMENTS

We would like to thank the NSF for supporting the project from which we collected the data for the evaluation (NSF National SMETE Digital Library: “Intelligence Collection Service for and about Educators and Students: Logging, Spidering, Analysis and Visualization,” Award No. DUE-0121741, Program 7444, September 2001 August 2003) and Dr. Edward Fox and Rao Shen, from the Virginia Polytechnic and State University. We would also like to thank the GetSmart team in the Artificial Intelligence Lab at the University of Arizona, who developed the concept mapping tools. Special thanks go to Dr. Hsinchun Chen and Mr. Byron Marshall for their support and advice in the development of this research and an early version of this paper.

REFERENCES