Knowledge Representation Structure for Cloud Platforms: SUMMUS Semantic Encoding Forest

S.S. Xie, Y. Tang, Z.X Miao, L. Sun, Y. He
School of Computer and Information Science
Southwest University
Chongqing, China

Abstract—Big data era, modern knowledge representation structures do not have appropriate structural support for parallel computing, unable to meet the time requirements of large data processing. We proposed one knowledge representation structure for cloud platform called SUMMUS Semantic Encoding Forest. Through semantic concept further fine-grained decomposition to achieve a unified representation of semantic relations and make binary semantic content-based encoding, then from structure level support parallel computing of semantic reasoning. This structure have better capabilities than the traditional knowledge represent structure at easy programming, complexity of rules and better structural etc. aspects. The time complexity of knowledge reasoning is limited to O (n). Experiments show that knowledge inference time of SUMMUS Semantic Encoding Forest do not obviously increased with the consumption of entities' or events' number of conflicts linearly increase, time performance is better than other existing knowledge representation structures.

Keywords-knowledge representation; big data; cloud platform; semantic encoding forest

I. INTRODUCTION

Currently, the mainstream knowledge representation like, the logical representation, the production rule, the semantic network, the frames and the scripts[1][2][3]. Existing knowledge representation corresponding set of rules and structures are lack of appropriate support to the distributed computing environment[1][4][5], cannot meet the time requirements of large data processing, so that the efficiency of knowledge reasoning is severely restricted. And with the knowledge base growth, the issue will be more obviously[2][6].

For enhance the reasoning efficiency of knowledge representation structure, proposed a parallel knowledge-based reasoning knowledge representation structure that appropriate the distributed feature of cloud platform, SUMMUS Semantic Encoding Forest. Through decomposition granularity control of the concept of semantic, to achieve a unified semantic relations. Proposed a knowledge organization tree based on entity classification, enhanced scalability of knowledge structure, facilitated knowledge relationship maintenance by encoding structure that proposed in this paper. Proposed semantic content-based binary encoding, reducing the conceptual structure of traditional knowledge structure itself brings knowledge reasoning complexity. Meanwhile, the encoding structure enables semantic reasoning to support parallel computing at the structure level, the time complexity of knowledge reasoning is limited to O(n).

II. SUMMUS SEMANTIC ENCODING FOREST

Definition 1: SUMMUS Semantic Encoding Forest, one structure for knowledge representation. knowledge broken down into a tree structure by according with its classification, and then encoding the content by the mapping rules of the structure. Ultimately represent knowledge and complete knowledge reasoning operating in coded form.

A. SUMMUS Encoding Forest Structure

Based on information management and systems of taxonomic knowledge decompose existing knowledge as classification tree. Parse entity into one of knowledge logic relationship structure in that knowledge that has been defined and independent of time will be decomposed into entities and relationships between them, represented and organized as a tree structure. Entities' Event is divided into four levels. Entity in the event tree root is a specific entity object. The second level of the hierarchy is the time level where every minute of the hierarchy date will correspond to a specific layer of the tree structure. The third level of the hierarchy is the geographic information level, the hierarchy in the encoding sequence as the place of the event's happened. The fourth level is the event description.

B. SUMMUS Encoding Mechanism

1) Encoding idea: SUMMUS semantic encoding taken content as base, using ANSI Chinese 2-byte fixed-length encoding into two types of encoding chains. One is the 1024 fixed-length encoding, and the other is elastic expansion encoding. The former is a major store and knowledge reasoning chain. While the latter only as details supplement of the corresponding field of the former. Why encoding more efficiency in reasoning is that the encoding structure reduces complexity of the concept of traditional knowledge representation structure by degradation in coded form. Meanwhile, the structure has good supportability to the parallel processing.

2) Encoding mapping structure: When encoding Chinese, semantic forest structure is mapped to the two kind of encoding structure as Figure 1 and Figure 2.

In 1024 fixed-length encoding chain, 0 to 63 bits for disambiguation flag and elastic expansion flag. Disambiguation flag is using to eliminate ambiguity at when the same position collision in content encoding. Elastic expansion flag indicates whether the corresponding bit segment requires flexible code.

Elastic expansion encoding chains are reserved the uniqueness part of the 1024 fixed-length encoding as the identify, than other parts using to expand the detailed   of the contents.
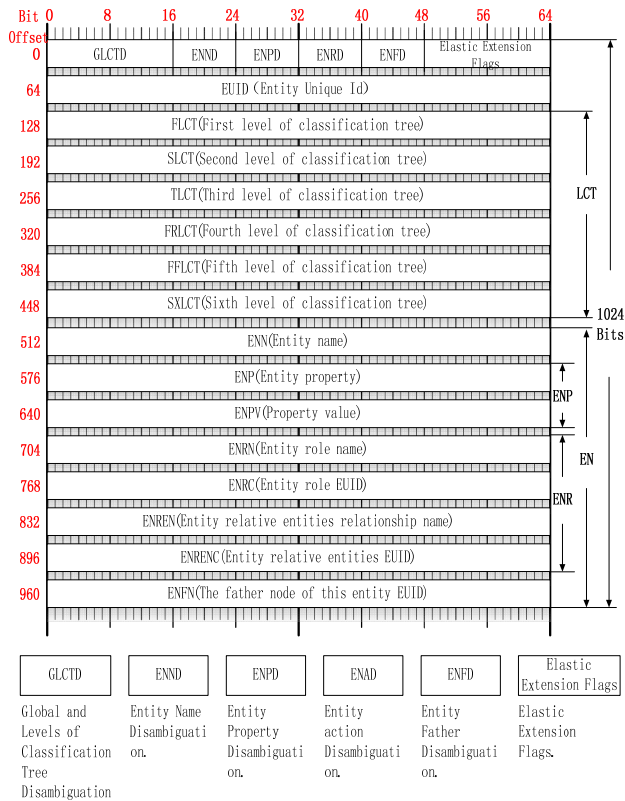

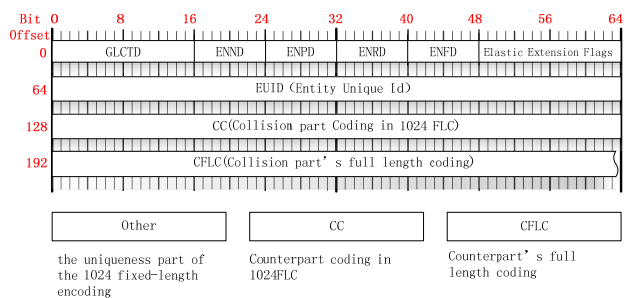FIGURE I. 1024 FIXED-LENGTH ENCODING CHAIN OF CLASSIFICATION ENTITY FOREST

FIGURE II. ELASTIC EXPANSION ENCODING CHAIN OF CLASSIFICATION ENTITY FOREST

III. KNOWLEDGE REASONING EFFICIENCY EXPERIMENT AND ANALYSIS

A. Experiment

Experiments based on the News corpus that provided by Sogou R&D center. And we only used 1900 news article in Technology category to constructed two prototype knowledge base systems one is based on the SUMMUS semantic forest, and another is knowledge represent structure Frames. We choose the reasoning goal is "Jake Ma's intention in September".

Time comparison points is 50 gradient increasing number of conflicts similar entities. After each increment of conflict, all the data randomly disrupted. We statistic reasoning time of each prototype system, under the same random order, same conflict entities number, and then,   we draw the contrast curves.

B. Result and Analysis

Currently, Frames is mainstream of knowledge representation structure, the existing knowledge base YAGO2, DBpedia etc. based on Frames. Knowledge Base Population issues of international frontier textual analysis meeting TAC are using Frames as the structure. Frames have better time performance than other existing knowledge represent structure of knowledge reasoning. Therefore, the experiment select Frames structure as the traditional knowledge structure representation to take part in comparative experiments of knowledge inference time. We use the "Traditional KR structure" as Frames' label.
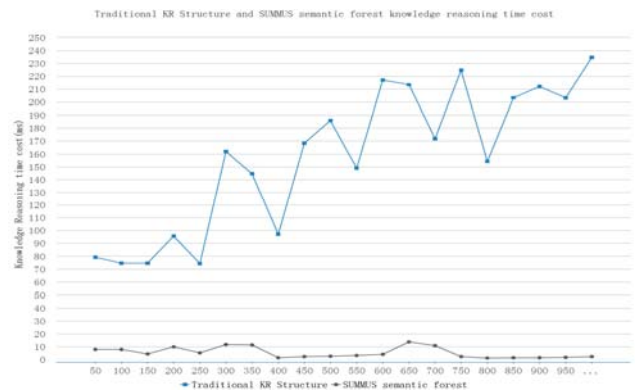

FIGURE III. KNOWLEDGE INFERENCE TIME CURVE WITH INCREASE AMBIGUOUS ENTITY CONFLICT

From the figure 3 , we can see that SUMMUS Semantic Encoding Forest can more effectively reasoning than traditional knowledge representation structure. knowledge inference time of SUMMUS semantic coding forest do not obviously increased with the consumption of entity or event   the number of conflicts linearly increase.

IV. SUMMARY

We proposed SUMMUS Semantic Encoding Forest, further fine-grained decomposition of semantic concept to achieve an unified representation of semantic relations and make binary semantic content-based encoding structure, then from structure support parallel computing of semantic reasoning. The time complexity of knowledge reasoning is limited to O (n) and do not linearly increased with the consumption of entities' or events' number of conflicts linearly increase, the time performance is better than the existing knowledge representation structures. Comparative show that this structure have better capabilities than the traditional knowledge represent structure at easy programming, the complexity of rules and better structural capability of concept.

The next step is to determine the optimal length of fixed-length encoding chain, design generic encoding mapping structure for multilingual, and implemented SUMMUS Semantic Encoding Forest in hardware to break through bottlenecks of knowledge inference time cost.

REFERENCES

[1] Tanwar, Poonam, T. V. Prasad, and S. Aswal. "Comparative Study of Three Declarative Knowledge Representation Techniques." International Journal on Computer Science & Engineering (2010).

[2] Furbach U. Representation and Reasoning[J]. Intelligent Systems, IEEE, 2014, 29(1): 2-2.

[3] Schank R C, Abelson R P. Scripts, plans, goals, and understanding: An inquiry into human knowledge structures[M]. Psychology Press, 2012.

[4] Sharma T, Kelkar D. A Tour Towards Knowledge Representation Techniques[J]. International Journal of Computer technology and Electrical Engineering (IJCTEE), 2011, 2(2): 2249-6342.

[5] Deshmukh P P, Ali M S. Overview of Xml based Knowledge Representation using Scripts[J]. International Journal, 2014, 2(1).

[6] Goertzel B, Pennachin C, Geisweiller N. Knowledge Representation Using the Atomspace[M]. Engineering General Intelligence, Part 2. Atlantis Press, 2014: 31-53.