Improvement of Schema-Informed XML Binary Encoding Using Schema Optimization Method

BumSuk Jang and Young-guk Ha'

Konkuk University, Department of Computer Science and Engineering,
Seoul, Republic of Korea
mimaster@naver.com and ygha@konkuk.ac.kr

Abstract. As XML documents are widely used for data communication on The Internet, there are many efforts to increase efficiency of XML data transmission. One of these efforts is to reduce the size of XML document such as BiM and EXI. These efforts are based on transformation of XML document into binary data. Basic principle of binary encoding is to assign XML event such as Element, Attribute, and Content to small code based on the schema. Size of the code of binary format is determined by the number of next XML events occurs. That is, size of schema affects size of binary encoded data. Generally, whole schema is used to encode the XML document, even though the XML document conforms to only a part of the schema. To reduce the size of binary encoded data more efficiently, this paper proposes an encoding method using a novel optimization technique for the XML schema used to encode XML documents.

1 Introduction

As XML documents are widely used for data communication on The Internet, there are many efforts to increase efficiency of XML data transmission. One of these efforts is to reduce the size of XML document such as BiM and EXI. Binary Mpeg7 (BiM) [1] is binary encoding format of Mpeg7 specification and The Efficient XML Interchange (EXI) format [2] is W3C's XML binary encoding standard. These efforts are based on transformation of XML document into binary data based on automata which derived from XML schema. State of automata means XML event and each state connected to XML event which will occurs in next step. Basic principle of XML binary encoding is that state transition was assigned to code. The code of state transition only depends on the number of transition from same state. Thus, length of the code is determined as log n where n is number of transition from same state. Size of schema affects size of binary encoded data because as size of schema increase, number of state of automata was increased also. Most of XML document over informed from schema despite it conforms to only a part of the schema. Therefore, to reduce the size of binary encoded data more, this paper proposes an encoding method using a novel optimization technique which extracts a part of the schema was conformed from the

[1] Corresponding author

adfa, p. 1, 2011.

XML document for the XML schema using EXI. The rest of this paper organized as follows. How schema optimization scheme reduces size of EXI encoded data was explained in section 2. In section 3, architecture of implementation was described. In section 4, show results of experiments. In section 6, conclusion and future work are described.

2 Schema Optimization Method

Basic principle of binary encoding is based on automata which derived from schema. Each state of automata derived from XML event. The transition to next state is assigned as a code. Size of the code is determined as log n where n is number of transition from same state. Thus, size of binary encoded data is increased by the number of automata transition. Most of XML document over informed from schema for conforms to only a part of the schema. Therefore, schema optimization technique which extracts a part of the schema was conformed from the XML document is efficient to reduce size of data. In additional, this method has less binary encoding time because has less the number of automata transition than original method.

Proposed encoding method using schema optimization technique is divided two parts. That is XML structure analyzing and optimized schema generation. For the communication between these functions, Fingerprint which represents XML document's structure is proposed. Fingerprint consists of bit flag which indicate usages of element was defined from schema. If XML document use the element, bit flag set to 1 and else set to 0. That is, Fingerprint represents part of original schema was matched with the XML.

There are two methods for schema optimization. One is Structure optimization method and the other is Definition optimization method. Structure optimization method is removes unused element in parent node's definition. On the other hand, Definition optimization method is removes unused element's definition. In section 4, we measure encoding performance improvement of schema optimization method using both of them.

3 Implementation of Schema Optimization Method

Module structure of original EXI encoding process consists of 2 modules. That is EXI grammar generator and EXI encoder. Inputs of EXI encoder are XML document and grammar which generated from schema conformed from the XML document by EXI grammar generator. There are two additional modules than original process for schema optimization technique insertion. Additional parts are XML structure analyzer (XSA) and Schema generator. XSA analyze XML structure and generate Fingerprint. Schema generator generate optimized schema using received Fingerprint. Generated optimized schema is used for EXI grammar generator.

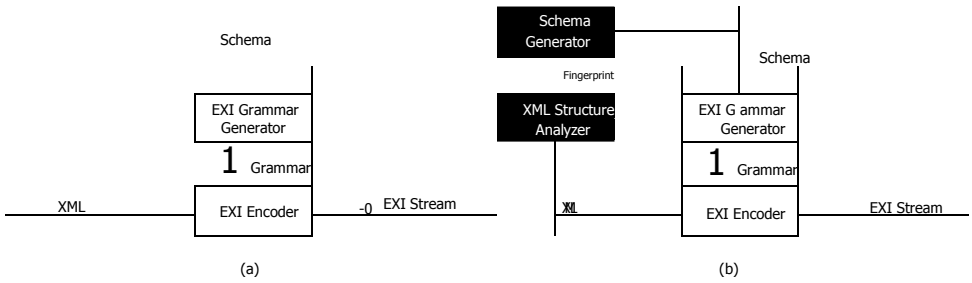

Fig. 1. Module structure of encoding process. (a) is module structure of original EXI encoding process. (b) is module structure of schema optimization method added EXI encoding process.

4 Experiments

TV-anytime Forum Metadata specifications [3] made by BBC was used for the experiment. Exifficient which developed by Siemens as open source project was used as EXI encoding engine. Fig. 2 Shows results of experiments. Graph (a) showed schema optimization technique based encoding method reduce size of encoded data than original encoding method. Graph (b) showed schema optimization technique based encoding method requires less encoding time than original encoding method.

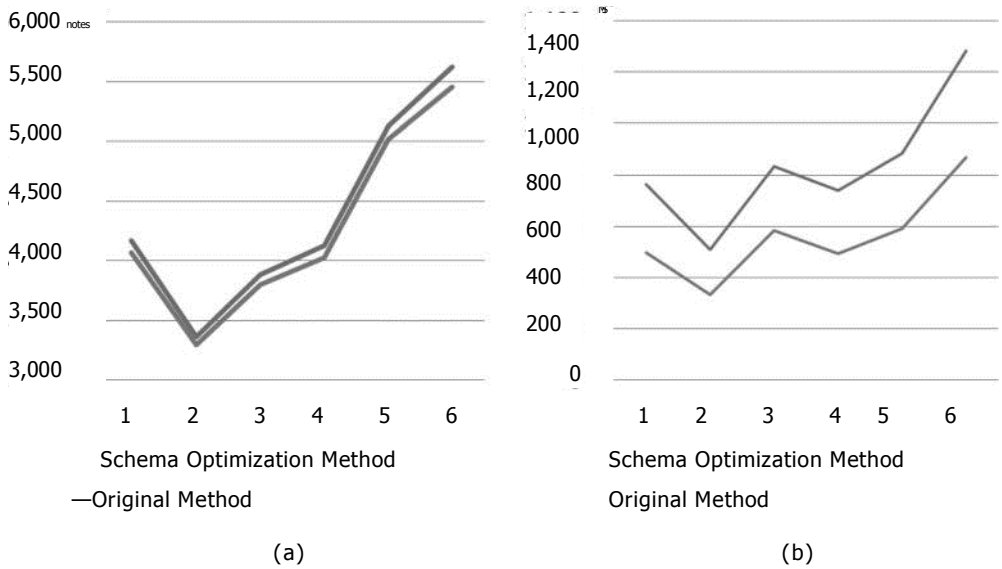

Fig. 2. Comparison of schema optimization technique based encoding method and original method. (a) is comparison of encoding size of schema optimization technique based encoding method with original encoding method. (b) is comparison of encoding speed of schema optimization technique based encoding method with original encoding method.

5 Conclusion and Future Work

This paper analyzed XML binary encoding methods and proposed method which was improved by schema optimization technique in schema informed XML binary encoding. Schema optimization technique was divided two modules which XML analyzer and Schema generator. Moreover, Fingerprint format that was defined for communication between these modules was proposed. Fingerprint is format which represents usage of the element of XML Schema was conformed from the XML document. As shown in the results of experiments, schema optimization technique based encoding method more efficient than original encoding method. We conclude that Schema optimization technique is useful to reduce size and time cost of XML binary encoding and it could improve the utility of bandwidth of the Internet. Another concern in this work, Schema generator was required for binary data decoding on client side. As mobile devices are used widely, client side issues are becomes important part of binary encoded XML data transmission. In that respect this technique required less computational cost because optimized schema which has less transition of automata. More concern in this work, motivation of schema optimization technique is reduce the transition of automata derived from schema. Because high level of depth in schema has only few elements, efficiency of schema optimization technique is depends on schema optimization level. In the future, we will define optimization level and discuss efficiency of each optimization level.

6 Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(grant number: 2012006817).

References

1. J. Heuer, C. Thienot, and M. Wollborn, MPEG-7 Binary Format.MPEG-7, 2002, ch. 2.3, Introduction to MPEG-7: Multimedia Content Description Interface.
2. J. Schneider and T. Kamiya, "Efficient XML Interchange (EXI) Format 1.0," http://www.w3.org/TR/exi/, W3C Candidate Recommendation 08 December 2009.
3. "The TV-Anytime Metadata specification", http://www.tv-anytime.org/workinggroups/wgmd.html