

# การบีบอัดข้อมูลเอ็กซ์เอ็มแอลโดยใช้วิธีจัดเรียงข้อมูลแบบกระชับ

## XML Data Compression using Brevity Encoding

ประพันธ์ เลขาโสภณ และ กานดา รุณนะพงศา

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น 40002

Tel./Fax : 0-4336-2160 E-mail : (superjing@gmail.com) , (krunapon@kku.ac.th)

---

**ABSTRACT** – XML becomes the standard language for data representation and exchange on the Internet because it is simple, flexible, and platform neutral. However, XML data is often large and verbose since it consists of many repeated tags which are used for self-describe data. The large size of data results in an excessive amount of space required for storing data on the disk space and time required for transmitting data over the network. Thus, it is necessary to find an effective compression technique for XML data. Previous compression techniques generate the compressed XML data that is in the binary format or in the proprietary format to only that XML compressor tool. Such format is often incomprehensible to others or difficult to automatically parse. In this work, we propose XBrevity, an XML compressor which supports compressing and uncompressing XML data. XBrevity adopts a novel encoding method that has the compressed XML data in the XML format. Thus, the compressed XML data still preserves the advantage features of XML but the XML document has the smaller size.

**KEY WORDS** – XML, Compression, Data exchange, Brevity

**บทคัดย่อ** – ภาษาเอ็กซ์เอ็มแอล (Extensible Markup Language, XML) ได้วิวัฒนาการมาเป็นภาษามาตรฐานในการเสนอและแลกเปลี่ยนข้อมูลบนอินเทอร์เน็ตเนื่องจากภาษา XML เป็นภาษาที่ใช้ง่ายมีความยืดหยุ่น และไม่ขึ้นอยู่กับระบบปฏิบัติการของเครื่องคอมพิวเตอร์ แต่ว่าข้อมูลที่เป็นภาษา XML นั้นมักจะมีขนาดใหญ่และมีข้อมูลที่ซ้ำซ้อนอันเนื่องมาจากการใช้แท็กที่ซ้ำ ๆ ในการอธิบายข้อมูล เอกสาร XML จึงมีขนาดใหญ่ซึ่งทำให้ต้องการพื้นที่ในการจัดเก็บในปริมาณมากและใช้เวลานานในการส่งข้อมูล ดังนั้นงานวิจัยที่ต้องการจะบีบอัดข้อมูลในภาษา XML จึงมีความจำเป็นอย่างยิ่ง งานวิจัยที่ได้ทำมาแล้วนั้น ข้อมูลที่ถูกบีบอัดแล้วอาจจะอยู่ในรูปแบบของไบนารีหรือรูปแบบที่เป็นที่เข้าใจกับผู้พัฒนาเครื่องมือเท่านั้น ในงานวิจัยนี้ได้นำเสนอวิธีบีบอัดข้อมูลแบบใหม่ชื่อว่า “เอ็กซ์เบรวิตี้ (XBrevity)” ซึ่งเป็นวิธีการที่ใช้ในการบีบอัดข้อมูลหรือคลายการบีบอัดข้อมูลโดยที่ข้อมูลที่ถูกบีบอัดสามารถเป็นที่เข้าใจได้กับคนทั่วไป การบีบอัดข้อมูลโดยที่เอกสารที่ถูกบีบอัดอยู่ในรูปแบบของภาษา XML แบบย่อนั้นจะทำให้เอกสารนั้นยังคงข้อดีของภาษา XML ในขณะที่ขนาดของเอกสารเล็กลงข้อมูลที่ถูกรบีบอัดอยู่ในรูปแบบของภาษา XML

**คำสำคัญ** – เอ็กซ์เอ็มแอล, การบีบอัดข้อมูล, การแลกเปลี่ยนข้อมูล, การย่อ

---