

ภาษาเอกซ์เอ็มแอล (Extensible Markup Language, XML) ในปัจจุบันได้วิวัฒนาการมาเป็นภาษามาตรฐานในการเสนอและแลกเปลี่ยนข้อมูลบนอินเทอร์เน็ต เนื่องจากภาษาเอกซ์เอ็มแอลเป็นภาษาที่ใช้งานง่ายมีความยืดหยุ่น และไม่ขึ้นอยู่กับระบบปฏิบัติการของเครื่องคอมพิวเตอร์ แต่ว่าข้อมูลที่เป็นภาษาเอกซ์เอ็มแอลนั้นมักจะมีขนาดใหญ่และมีข้อมูลที่ซ้ำซ้อนอันเนื่องมาจากการใช้แท็กที่ซ้ำๆ กันในการอธิบายข้อมูล เอกสารเอกซ์เอ็มแอลจึงมีขนาดใหญ่ซึ่งทำให้ต้องการพื้นที่ในการจัดเก็บในปริมาณมากและใช้เวลานานในการส่งข้อมูล ดังนั้นงานวิจัยที่ต้องการจะบีบอัดข้อมูลในภาษาเอกซ์เอ็มแอลจึงมีความจำเป็นอย่างยิ่ง ในงานวิจัยที่ได้ทำมาแล้วนั้น ข้อมูลที่ถูกบีบอัดแล้วอาจจะอยู่ในรูปแบบของไบนารีหรือรูปแบบที่เฉพาะผู้พัฒนาเครื่องมือเท่านั้นที่เข้าใจ ในงานวิจัยนี้ได้นำเสนอวิธีบีบอัดข้อมูลแบบใหม่ชื่อว่า “เอกซ์เบรวิตี้ (XBrevity)” ซึ่งเป็นวิธีการที่ใช้ในการบีบอัดข้อมูลหรือคลายการบีบอัดข้อมูลโดยที่ข้อมูลที่ถูกบีบอัดสามารถเป็นที่เข้าใจได้ง่ายกับคนทั่วไป โดยที่วิธีบีบอัดที่นำเสนอนี้สามารถใช้ได้กับเอกสารเอกซ์เอ็มแอลใดๆ รวมทั้ง สามารถใช้กับเอกสารเอกซ์เอ็มแอลที่ไม่มีเอกสารที่อธิบายโครงสร้างของเอกสารเอกซ์เอ็มแอล (เอกสาร XML Schema)

งานวิจัยนี้เสนอแนวทางการบีบอัดข้อมูลโดยที่เอกสารที่ถูกบีบอัดอยู่ในรูปแบบของภาษาเอกซ์เอ็มแอลแบบย่อซึ่งจะทำให้เอกสารนั้นยังคงข้อดีของภาษาเอกซ์เอ็มแอล ขอบเขตของงานวิจัยนี้ไม่ได้มีเป้าหมายในการบีบอัดเอกสารเอกซ์เอ็มแอลให้ขนาดของเอกสารเอกซ์เอ็มแอลมีขนาดเล็กที่สุดเมื่อเปรียบเทียบกับวิธีการบีบอัดข้อมูลอื่นๆ แต่งานวิจัยนี้มีจุดประสงค์ให้ข้อมูลของเอกสารเอกซ์เอ็มแอลที่ถูกบีบอัดแล้วเป็นที่เข้าใจง่ายและยังอยู่ในรูปแบบของเอกสารเอกซ์เอ็มแอล วิธีการวิจัยเริ่มจากการออกแบบโครงสร้างของเอกสารเอกซ์เอ็มแอลในรูปแบบย่อ ออกแบบอัลกอริทึมที่ใช้ในโปรแกรมที่ใช้ในการบีบอัดและการคลายการบีบอัด จากนั้นทำการพัฒนาโปรแกรม และทำการวัดประสิทธิภาพของการบีบอัดตลอดจนเวลาที่ใช้ในการบีบอัด โดยใช้ข้อมูล เอกซ์เอ็มแอล จากการดาวน์โหลดไฟล์ เอกซ์เอ็มแอล จากอินเทอร์เน็ตและใช้โปรแกรม XMark ในการสร้างไฟล์เอกซ์เอ็มแอลแบบสุ่ม จากผลการทดลองพบว่าโปรแกรมบีบอัดข้อมูลสามารถทำให้ข้อมูลเล็กลง และได้ข้อมูลที่อยู่ในรูปแบบเอกซ์เอ็มแอล

eXtensible Markup Language (XML) becomes the standard language for data representation and exchange on the Internet because it is simple, flexible, and platform independent. However, XML data is often large and verbose since it consists of many repeated tags which are used to 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. In previous compression techniques, the compressed data is in the binary form or in the format that is understandable for only a few researchers who own those techniques. 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 format that is easy to understand for anyone. This method can apply for any XML document including one without its associated XML Schema file defined.

This research work proposes the method in compressing the original XML data into the compressed XML data using a brevity format. Thus, the compressed XML data still preserves the advantage features of XML but the XML document has a smaller size. The scope of this research work does not aim to minimize the size of the compressed document to be the smallest among all compressing methods. The goal of this research work is rather to make the compressed data easy to understand and still be in the XML format. The research method consists of designing the brevity form of compressed XML document, designing the algorithms in compressing and decompressing XML data, developing the programs, and measuring the effectiveness and time in compressing XML data. To test the compression and decompression techniques, we use input files which are downloaded from the Internet and the random generated files using XMark. Based on the experimental results, we found that the compressed data is in the XML file format and has the smaller size.