เทคโนโลยีเว็บและมาตรฐานสำหรับห้องสมุดดิจิทัล

(Web Technology and Standards for Digital Libraries)

อรนุช เศวตรัตนเสถียร*

บทนำ

ปัจจุบัน เทคโนโลยีเว็บ (Web technology) มีความก้าวหน้าเป็นอย่างมาก และมีผลกระทบอย่างกว้างขวาง เช่น ผลกระทบต่อการดำเนินธุรกิจ การศึกษา หรือการประชาสัมพันธ์ รวมถึงการให้บริการสารนิเทศขององค์กรประเภทห้องสมุด โดยเฉพาะอย่างยิ่งเมื่อมีแนวคิดเกี่ยวกับห้องสมุดดิจิทัล (digital library) เกิดขึ้น ดังจะเห็นได้ว่า โครงการห้องสมุดดิจิทัลต่าง ๆ ที่ดำเนินการอยู่นั้นให้บริการด้วยการเข้าถึงทางเวิลด์ไวด์เว็บ (World Wide Web - WWW)

เนื่องจากห้องสมุดดิจิทัลมีความสัมพันธ์กับเทคโนโลยีเว็บและมาตรฐานต่าง ๆ ความรู้เกี่ยวกับประเด็นที่เกี่ยวข้องนี้ จึงมีส่วนช่วยให้เกิดความเข้าใจมากยิ่งขึ้นเพื่อการดำเนินงานห้องสมุดดิจิทัลได้ บทความนี้ จะนำเสนอสารนิเทศเกี่ยวกับแนวคิดเรื่องห้องสมุดดิจิทัล เวิลด์ไวด์เว็บ และมาตรฐานพอสังเขป เพื่อความเข้าใจเบื้องต้นในประเด็นเหล่านี้ แล้วจึงนำเสนอมาตรฐานต่าง ๆ ที่เกี่ยวข้องในแต่ละด้าน ได้แก่ การสื่อสาร การเข้ารหัส (encoding) การแสดงรายละเอียดทรัพยากร (resource description) และสื่อแต่ละประเภท ตามลำดับ

ห้องสมุดดิจิทัล

"ห้องสมุดดิจิทัล" คงเป็นคำคุ้นเคยที่ได้ยินกันค่อนข้างแพร่หลายในวงการบรรณารักษศาสตร์และสารนิเทศศาสตร์ในขณะนี้ อย่างไรก็ดี หากกล่าวถึงนิยามของคำดังกล่าว หลายคนอาจจะไม่แน่ใจหรือรู้สึกว่านิยามของคำดังกล่าวตามความเข้าใจยังไม่ชัดเจนพอ ในที่นี้จึงขอนำเสนอนิยามของคำว่า "ห้องสมุดดิจิทัล" จาก Digital Library Federation (DLF) ดังนี้

---

* อาจารย์ประจำภาควิชาบรรณารักษศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์-มหาวิทยาลัย

ห้องสมุดดิจิทัล หมายถึง องค์กรที่มีทรัพยากรต่าง ๆ ให้ อันรวมถึงบุคลากรที่มีความเชี่ยวชาญ เพื่อคัดเลือก จัดโครงสร้าง ให้การเข้าถึงทางปัญญา ตีความ เผยแพร่ สงวนรักษาความเป็นบูรณาการ และให้ความมั่นใจได้ถึงความคงอยู่ (persistence) ผ่านกาลเวลาของทรัพยากรงานดิจิทัล เพื่อว่าทรัพยากรต่าง ๆ นั้นจะมีพร้อมให้ใช้ได้อย่างประหยัดสำหรับการนำไปใช้โดยชุมชนใด ๆ หรือกลุ่มชุมชนที่ระบุได้ (Waters 1998)

จากนิยามดังกล่าว มีประเด็นที่น่าสนใจว่า "ห้องสมุดดิจิทัล" มิได้หมายถึงเพียง "ทรัพยากรต่าง ๆ" ตามที่อาจเคยเข้าใจกันเมื่อมีการกล่าวถึงเวิลด์ไวด์เว็บในฐานะห้องสมุดดิจิทัล หากแต่หมายถึง "องค์กร" ที่รองรับทรัพยากรเหล่านั้น ทั้งนี้ห้องสมุดดิจิทัลมีหน้าที่ต่าง ๆ มากมาย และเมื่อพิจารณา จะพบว่า หน้าที่ที่ดำเนินการได้ยากยิ่งเนื่องด้วยความเปลี่ยนแปลงที่มีอยู่ตลอดเวลา แต่เป็นหน้าที่ที่มีความสำคัญต่อแนวคิดเรื่องห้องสมุดดิจิทัล คือ การสงวนรักษาความเป็นบูรณาการ และการให้ความมั่นใจได้ถึงความคงอยู่แม้กาลเวลาจะล่วงเลยไปของทรัพยากรดิจิทัล ในการดำเนินงานจริง จึงไม่ควรละเลยประเด็นดังกล่าว สำหรับทรัพยากรของห้องสมุดดิจิทัลหลายคนอาจเข้าใจว่าคำนึงจากรูปแบบของทรัพยากรที่เป็นดิจิทัลเท่านั้น ซึ่งไม่ถูกต้องนัก เพราะหลักที่ใช้กำหนดนโยบายด้านทรัพยากร ควรพิจารณาจากเนื้อหาของทรัพยากร และชุมชนผู้ใช้ทรัพยากร โดยความต้องการและความสนใจของชุมชนผู้ใช้ที่ระบุได้นี้ เป็นสิ่งที่กำหนดแนวทางการพัฒนาของห้องสมุดดิจิทัล อันรวมถึงการลงทุนที่ห้องสมุดดำเนินการ ทั้งในด้านเนื้อหาและเทคโนโลยีด้วย นอกจากนี้ ห้องสมุดดิจิทัลยังต้องตระหนักถึงความสำคัญของคุณภาพการบริการ เพราะห้องสมุดดิจิทัลเป็นองค์กรที่ดำเนินงานให้บริการ ประเด็นสำคัญที่ใช้ในการพิจารณานี้ได้แก่ ความพร้อมรวมถึงการตอบรับในการมีสารนิเทศให้ผู้ใช้ของห้องสมุดใช้ได้ และค่าใช้จ่าย โดยห้องสมุดดิจิทัลที่ประสบความสำเร็จควรเข้าใจและดำเนินงานอย่างรวดเร็วเพื่อลดปัจจัยที่เป็นค่าใช้จ่ายได้

เวิลด์ไวด์เว็บ

เวิลด์ไวด์เว็บ ซึ่งบางครั้งเรียกกันว่า เว็บสเปซ (webspace) ไซเบอร์สเปซ (cyberspace) หรือเว็บ (Web) นั้น หมายถึง ระบบสารนิเทศที่เป็นเครือข่ายแบบดิจิทัล ซึ่งไม่ใช่อินเทอร์เน็ต (Internet) แต่เป็นบริการที่ใช้อินเทอร์เน็ตและเครือข่ายต่าง ๆ เป็นตัวกลางในการสื่อสารข้อมูล เวิลด์ไวด์เว็บช่วยให้สามารถอ่านแฟ้มข้อมูลและเอกสารอิเล็กทรอนิกส์ต่าง ๆ จากทั่วโลกได้ด้วยการเชื่อมโยงของข้อความหลายมิติ (hypertext link) โดยเวิลด์ไวด์เว็บให้บริการการสื่อสารข้อมูล

บนเครือข่ายอินเทอร์เน็ตหลายรูปแบบ ทั้งข่าวสาร ข้อมูลที่เป็นตัวอักษร ภาพกราฟิกส์ และสื่อประสม (multimedia) รวมทั้งความสามารถในการโต้ตอบระหว่างผู้ใช้กับผู้ให้บริการ การเรียกข้อมูลจากเวิลด์ไวด์เว็บขึ้นมาดู ต้องใช้โปรแกรมที่เรียกว่า "เว็บเบราว์เซอร์" (web browser) เวิลด์ไวด์เว็บสามารถทำงานได้ในเครื่องคอมพิวเตอร์และโปรแกรมปฏิบัติการทุกชนิด (cross platform) โดยลักษณะการนำเสนอข้อมูลเป็นแบบ Graphical User Interface (GUI)* จึงสะดวกต่อการใช้มาก (ต้น ตัณฑ์สุทธิวงศ์, สุพจน์ ปุณณชัยยะ, และสุวัฒน์ ปุณณะชัยยะ 2539: 335; พรทิพย์ โล่ห์เลขา 2540: 4; สวัสดิ์ ไกรคุ้ม 2541: 17; Garlock and Piontek 1996: 2; Poulter, Tseng, and Sargent 1999:4)

เทคโนโลยีเว็บมีความสัมพันธ์กับห้องสมุดดิจิทัล ดังที่มีคำกล่าวว่า "...เวิลด์ไวด์เว็บและเทคโนโลยีที่เกี่ยวข้องมีความสำคัญต่อการเจริญเติบโตอย่างรวดเร็วของห้องสมุดดิจิทัล..." (Arms 2000: 29) ทั้งนี้ อาจเนื่องจากเหตุผลต่าง ๆ ดังนี้

- เวิลด์ไวด์เว็บสามารถเชื่อมโยงเอกสารต่าง ๆ ได้ทั่วโลก และเชื่อมโยงไปยังข้อมูลของบริการอื่น ๆ บนอินเทอร์เน็ตได้ เช่น เทลเน็ต (Terminal Emulation Link Network - Telnet) เอฟทีพี (File Transfer Protocol - FTP) และยูสเน็ต (User Network - USENET)
- เวิลด์ไวด์เว็บแสดงข้อมูลได้หลายรูปแบบ ทำให้เกิดความน่าสนใจและความสนุกสนานเพลิดเพลิน
- เวิลด์ไวด์เว็บสามารถใช้กับคอมพิวเตอร์ระบบใด ๆ ก็ได้
- เวิลด์ไวด์เว็บใช้เผยแพร่เอกสารที่จัดทำขึ้นได้ทั่วโลกตลอด 24 ชั่วโมง และปรับปรุงข้อมูลให้ทันสมัยได้ตลอดเวลา
- ซอฟต์แวร์ที่ใช้บนเวิลด์ไวด์เว็บมีประเภทที่สามารถนำมาใช้ได้โดยไม่เสียค่าใช้จ่าย เช่น ซอฟต์แวร์สำหรับเครื่องบริการเว็บ (web server) ชื่อ Apache

สำหรับในส่วนนี้ จะขอนำเสนอเฉพาะความรู้เบื้องต้นเกี่ยวกับเวิลด์ไวด์เว็บเพื่อความเข้าใจในเทคโนโลยีดังกล่าวที่มีความเกี่ยวข้องกับห้องสมุดดิจิทัลดังนี้

โครงสร้างของเวิลด์ไวด์เว็บ

เมื่อกล่าวถึงเวิลด์ไวด์เว็บ คำศัพท์ที่เกี่ยวข้องซึ่งได้ยินกันเสมอ จะมีทั้งเว็บไซต์ (web site) เว็บเพจ (web page) และโฮมเพจ (home page) หลายคนอาจยังสับสน และใช้

* Graphical User Interface (GUI) หมายถึง ซอฟต์แวร์ที่ซ่อนส่วนที่เป็นเทคนิคไว้ โดยช่วยให้การใช้เครื่องคอมพิวเตอร์นั้น เป็นเพียงการใช้เมนูหรือการคลิกบนสัญรูป (icon)

คำต่าง ๆ เหล่านี้แทนกัน จึงขอกล่าวถึงคำศัพท์ทั้ง 3 คำนี้ เพื่อแสดงให้เห็นความสัมพันธ์ ซึ่งอาจจะช่วยให้เกิดความเข้าใจได้มากขึ้น

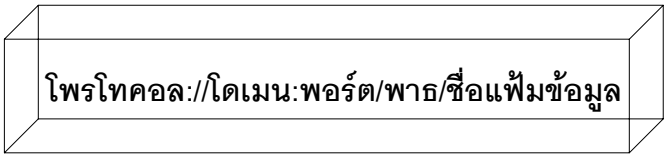
เวิลด์ไวด์เว็บประกอบด้วยเว็บไซต์ต่าง ๆ เป็นจำนวนมาก โดยเว็บไซต์ หมายถึง ชุดของเว็บเพจจำนวนหลาย ๆ หน้า ที่มีความสัมพันธ์กัน และมักได้รับการจัดระเบียบไว้บนเครื่องบริการ (server) เดียวกัน แต่หากเป็นเว็บไซต์ที่มีขนาดใหญ่ อาจใช้เครื่องบริการหลายเครื่องในทางกลับกันเครื่องบริการหนึ่ง อาจจัดเก็บเว็บไซต์หลายแห่งก็เป็นได้ เว็บไซต์ที่พบเห็นกันอย่างแพร่หลาย อาจสร้างขึ้นโดยห้องสมุด องค์กรของรัฐ องค์กรธุรกิจ กลุ่มบุคคล หรือบุคคล ส่วนเว็บเพจ หมายถึง เอกสารหน้าหนึ่ง ๆ บนเวิลด์ไวด์เว็บ ซึ่งให้รายละเอียดและเนื้อหาต่าง ๆ บนเว็บไซต์ ดังนั้น จึงอาจกล่าวโดยสรุปเชิงเปรียบเทียบได้ว่า เว็บไซต์เปรียบเสมือนหนังสือเล่มหนึ่ง ๆ ในขณะที่เว็บเพจเปรียบเสมือนหน้าหนังสือที่ประกอบกันเป็นหนังสือ

สำหรับโฮมเพจนั้น นอกจากจะหมายถึงเว็บไซต์ส่วนบุคคล ในกรณีที่เว็บไซต์ที่สร้างขึ้นเป็นเว็บไซต์ของบุคคลใด ๆ เป็นการส่วนตัว ตามที่ได้ยินกันโดยทั่วไปแล้ว โฮมเพจยังมีความหมายต่าง ๆ ดังนี้ (นิรุธ อำนวยศิลป์ 2542: 87; ชัยยุทธ์ ลิมลาวัลย์ 2544: 19; สวัสดิ์ ไกรคุ้ม 2541: 25; Metz and Junion-Metz 1996: 10-13; Arms 2000: 37)

- เอกสารที่แสดงผลเป็นเอกสารหน้าแรกของเว็บไซต์ เช่น index.html หรือ default.htm ตามความหมายนี้ โฮมเพจจึงอาจเปรียบได้กับหน้าปกของหนังสือ ซึ่งโฮมเพจนี้ ต้องดำเนินการเป็นอย่างดี มีการวางแผนและออกแบบให้แสดงผลได้อย่างรวดเร็ว สวยงาม และน่าติดตามชม เพื่อผู้ใช้จะได้เยี่ยมชมเว็บเพจอื่น ๆ ต่อไป
- เอกสารที่แสดงผลเป็นเอกสารหน้าแรกเมื่อเริ่มใช้งานเว็บเบราว์เซอร์ โดยอาจกำหนดด้วยโปรแกรมเบราว์เซอร์ หรือผู้ใช้เป็นผู้กำหนดเอง กรณีที่ผู้ใช้เป็นผู้กำหนดโฮมเพจตามความหมายนี้ด้วยตนเอง โฮมเพจอาจเป็นเว็บเพจที่มีสารนิเทศที่ผู้ใช้สนใจ หรือมีจุดเชื่อมโยงไปยังเว็บไซต์ต่าง ๆ ที่ใช้งานบ่อย
- ชุดสารนิเทศย่อยภายในเว็บไซต์ กล่าวคือ สารนิเทศที่เกี่ยวข้องกับแผนงาน โครงการ หรือบริการใด ๆ อาจมีการจัดทำโฮมเพจเฉพาะขึ้นภายในเว็บไซต์ของบริษัทหรือสถาบันนั้น ๆ

ที่อยู่ของเวิลด์ไวด์เว็บ

ในการใช้เวิลด์ไวด์เว็บ ยูอาร์แอล (Uniform/Universal Resource Locator - URL) เป็นกลไกที่ใช้ระบุที่อยู่ของเอกสาร และช่วยให้เกิดการเชื่อมโยงสารนิเทศจากที่ต่าง ๆ ได้ ยูอาร์แอลมีส่วนประกอบสำคัญดังนี้



โพรโทคอล (protocol)	โพรโทคอลบนอินเทอร์เน็ต ซึ่งถ้าเป็นเวิลด์ไวด์เว็บ จะใช้ http
โดเมน (domain)	ตำแหน่งของเครื่องคอมพิวเตอร์บนเครือข่ายที่ต้องการใช้บริการ
พอร์ต (port)	การเชื่อมต่อพิเศษกับเครื่องให้บริการบางแห่ง ซึ่งไม่ค่อยพบ
พาธ (path)	ไดเร็กทอรี (directory) ที่เก็บเอกสารที่ต้องการ
ชื่อแฟ้มข้อมูล (filename)	ชื่อแฟ้มข้อมูล (เอกสาร) ที่ต้องการ ซึ่งหากไม่ระบุ จะเป็นแฟ้มข้อมูลชื่อ index.html หรือ default.htm เป็นต้น

ตัวอย่าง

http://www.clir.org/pubs/issues/issues04.html

โพรโทคอล โดเมน พาธ ชื่อแฟ้มข้อมูล

อนึ่ง ชื่อโดเมนที่พบเห็นกันนั้น แบ่งออกได้เป็น 3 ระดับ คือ โดเมนระดับสูงสุด (Top Level Domain) โดเมนระดับที่สอง (Second Level Domain) และโดเมนระดับอื่น ๆ (Third Level Domain, ...) (ชัยยุทธ์ ลิมลาวัลย์ 2544: 29)

- โดเมนระดับสูงสุด เป็นระดับของชื่อโดเมนลำดับแรก เช่น .com .net .org .de .th .uk โดเมนระดับสูงสุดนี้ ยังแบ่งออกเป็น 2 ประเภท คือ โดมนระดับสูงสุดแบบสากล (General Internet DNS Top Level Domain - gTLD) ซึ่งเป็นชื่อโดเมนระดับสูงสุดที่ทุกประเทศสามารถจดทะเบียนได้ ยกเว้น .edu .gov .int และ .mil ที่ใช้ในประเทศสหรัฐอเมริกาเท่านั้น ทั้งนี้ เมื่อพฤศจิกายน พ.ศ. 2543 (ค.ศ. 2000) ได้มีการประกาศชื่อโดเมนสากลเพิ่มเติมอีก 7 ประเภทด้วย ได้แก่ .aero .biz .coop .info .museum .name และ .pro (พิฐุมา

พันธุ์ทวี 2543: 17-18) ส่วนโดเมนระดับสูงสุดอีกประเภท คือ โดเมนระดับสูงสุดที่เป็นรหัสประเทศ (Country Code Top Level Domain - ccTLD) ซึ่งเป็นชื่อโดเมนระดับสูงสุดที่แต่ละประเทศเป็นผู้บริหารดูแล เช่น .jp (ญี่ปุ่น) .th (ไทย) และ .uk (สหราชอาณาจักรอังกฤษ)

- โดเมนระดับที่สอง เป็นระดับของชื่อโดเมนลำดับที่สอง ซึ่งอาจเป็นชื่อของโดเมนในโดเมนระดับสูงสุดแบบสากล เช่น google.com standford.edu และ loc.gov หรือเป็นประเภทของโดเมนในโดเมนระดับสูงสุดที่เป็นรหัสประเทศ เช่น co.jp or.th และ ac.uk
- โดเมนระดับอื่น ๆ เป็นระดับของชื่อโดเมนลำดับที่สาม หรือลำดับที่อยู่ถัด ๆ ไปซึ่งมีการกำหนดขึ้น เช่น uwtc.washington.edu jsc.nasa.gov nectec.or.th และ dis.strath.ac.uk

มาตรฐาน

The International Organization for Standards (ISO) ให้นิยามของคำว่า "มาตรฐาน" ไว้ว่า มาตรฐาน เป็นเอกสารที่มีให้ใช้ได้ต่อสาธารณะ จัดทำขึ้นด้วยความร่วมมือและความเห็นพ้องต้องกันหรือการยอมรับของผู้ที่มีผลประโยชน์ ซึ่งได้รับผลกระทบ โดยมีพื้นฐานจากผลลัพธ์ที่รวบรวมขึ้นทางวิทยาศาสตร์ เทคโนโลยี และประสบการณ์ อันมีความมุ่งหมายเพื่อส่งเสริมผลประโยชน์สูงสุดของชุมชน และได้รับการรับรองจากองค์กรซึ่งเป็นที่รู้จักในระดับชาติ ภูมิภาค หรือนานาชาติ (Boss 2000: 7)

มาตรฐานเป็นสิ่งสำคัญที่ควรพิจารณาในการปฏิบัติการใด ๆ เพราะมาตรฐานเป็นสิ่งที่ช่วยผสมผสานความหลากหลายให้เกิดความเข้ากันได้ (compatibility) อันทำให้เกิดความสามารถในการทำงานระหว่างกัน (interoperability) อาจกล่าวได้ว่า การดำเนินงานที่สอดคล้องกับมาตรฐานช่วยให้เกิดความมั่นใจได้ว่าระบบที่จัดทำขึ้น จะสามารถเชื่อมโยงกับระบบอื่น ๆ ได้ นอกจากนี้มาตรฐานเป็นสิ่งที่เปลี่ยนแปลงค่อนข้างช้า โดยเฉพาะมาตรฐานที่เป็นทางการในระดับชาติและนานาชาติ อีกทั้งในกรณีที่มีการปรับเปลี่ยน โดยปกติแล้ว มาตรฐานฉบับใหม่กว่า ก็มักจะเป็นไปตามฉบับก่อนหน้า (Noerr 2000: 114) ดังนั้น มาตรฐานจึงเป็นสิ่งที่ช่วยให้เกิดความมั่นใจได้ด้วยว่าการลงทุนในระบบที่จัดทำนั้น จะไม่สูญเปล่าไปอย่างรวดเร็ว

มาตรฐานที่ใช้กันอยู่ อาจแบ่งออกได้เป็นมาตรฐานซึ่งมีเจ้าของสิทธิ (proprietary standard) พัฒนาและประกาศใช้โดยบริษัทต่าง ๆ เพื่อประกันหรือเพิ่มส่วนแบ่งทางการตลาด และมาตรฐานเปิด (open standard) ซึ่งเผยแพร่และมีให้ใช้ได้โดยบุคคลใด ๆ ก็ตาม ทั้งนี้ มาตรฐานทั้ง 2 ประเภท อาจเป็นมาตรฐานแบบนิตินัย (de jure standard) คือ มาตรฐานที่มีการรับรองโดย

องค์กรมาตรฐานอย่างเป็นทางการ หรือเป็นมาตรฐานแบบพฤตินัย (de facto standard) คือ มาตรฐานที่ใช้กันอย่างแพร่หลายจนกลายเป็นปกติ (Haigh 2001) มาตรฐานทั้ง 2 ประเภทนี้ จะเป็นมาตรฐานที่น่าสนใจสำหรับห้องสมุดดิจิทัลที่จะกล่าวถึงในลำดับถัดไป

มาตรฐานด้านการสื่อสาร

มาตรฐานด้านการสื่อสารที่จะกล่าวถึงในที่นี้ ประกอบด้วย มาตรฐานทีซีพี / ไอพี เอชทีทีพี ไมม์ Z39.50, ISO23950 และISO

ทีซีพี / ไอพี *(Transmission Control Protocol / Internet Protocol - TCP / IP)*

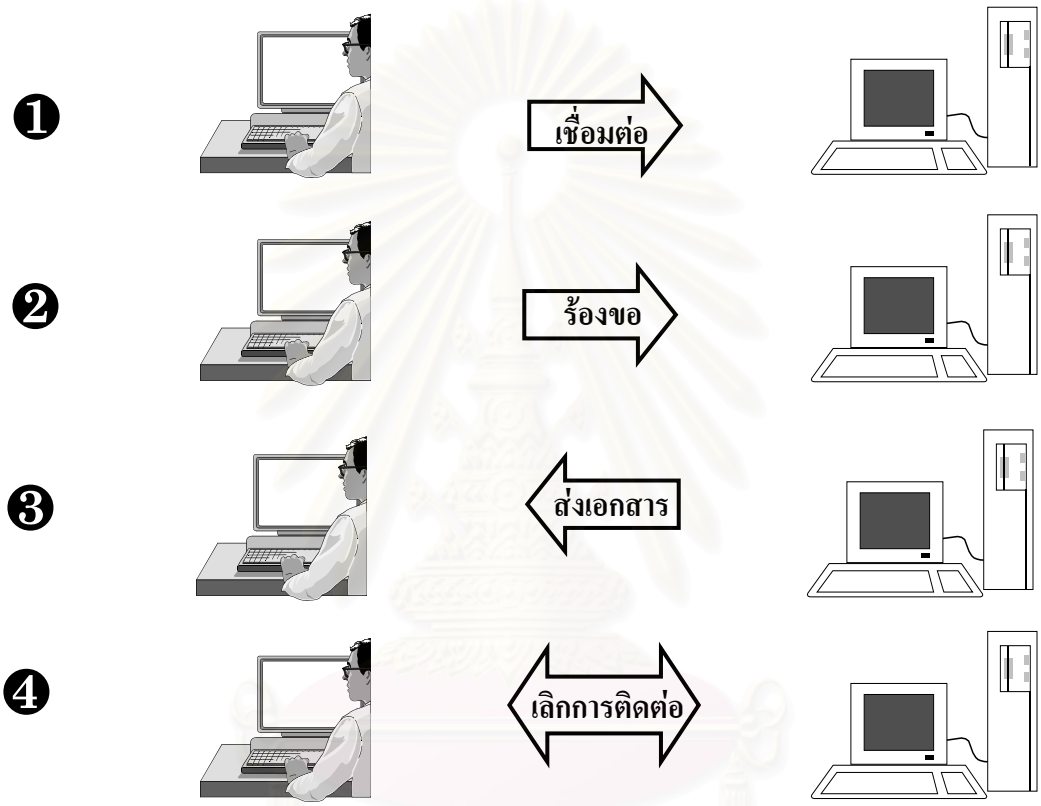
ทีซีพี / ไอพี เป็นโพรโทคอลที่ใช้เป็นมาตรฐานในการเชื่อมต่อสื่อสารระหว่างระบบคอมพิวเตอร์ ซึ่งเป็นที่รู้จักกันโดยทั่วไป เพราะเป็นโพรโทคอลที่ใช้กันบนเครือข่ายอินเทอร์เน็ต ทีซีพี / ไอพี ประกอบด้วยโพรโทคอลต่าง ๆ มากกว่า 100 โพรโทคอล โดยโพรโทคอลที่สำคัญนั้น เช่น เทลเน็ต ซึ่งอนุญาตให้ผู้ใช้จากที่ห่างไกล ล็อก (log) ไปยังแม่ข่าย (host) อื่น ๆ ได้ เอฟทีพี ซึ่งช่วยให้สามารถถ่ายโอนแฟ้มข้อมูลระหว่างเครื่องคอมพิวเตอร์บนอินเทอร์เน็ต และเอสเอ็มทีพี (Simple Mail Transfer Protocol - SMTP) ซึ่งเป็นโพรโทคอลที่ใช้สำหรับไปรษณีย์อิเล็กทรอนิกส์ หรืออีเมล (electronic mail - E-mail)

แม้ว่าทีซีพี / ไอพี จะมีกำเนิดจากกระทรวงกลาโหมของสหรัฐฯ (U.S. Department of Defense) แต่ในปัจจุบัน ผู้ที่รับผิดชอบดูแลมาตรฐานนี้ คือ International Engineering Task Force (IETF)

เอชทีทีพี (HyperText Transfer Protocol - HTTP)

เอชทีทีพี เป็นโพรโทคอลของเวิลด์ไวด์เว็บ ใช้เป็นมาตรฐานในการรับ - ส่งข้อมูลระหว่างเครื่องบริการเว็บและเครื่องรับบริการ (client) หรือเว็บเบราว์เซอร์ โดยเครื่องบริการเว็บเป็นเครื่องคอมพิวเตอร์ที่จัดเก็บเอกสารเว็บและสารนิเทศต่าง ๆ รับการร้องขอ ประมวลผล และส่งเอกสารต่าง ๆ ไปให้เว็บเบราว์เซอร์ในเครื่องรับบริการ และสามารถทำงานแบบโต้ตอบกับเว็บเบราว์เซอร์ได้ ตัวอย่างโปรแกรมที่ใช้สำหรับเครื่องบริการ เช่น Apache และ Internet Information Server (IIS) ส่วนเว็บเบราว์เซอร์ คือ โปรแกรมที่อยู่ในเครื่องรับบริการ ใช้สำหรับการเข้าถึงเพื่อเรียกดูเอกสารจากบริการเวิลด์ไวด์เว็บบนอินเทอร์เน็ต โดยสามารถแสดงผลเว็บเพจบนจอภาพเป็นรูปแบบต่าง ๆ ตัวอย่างเว็บเบราว์เซอร์ เช่น Internet Explorer (IE) Lynx Netscape Navigator / Communicator และ Opera

กระบวนการรับส่งข้อมูลด้วยเอชทีทีพี อาจสรุปได้เป็น 4 ขั้นตอน คือ เครื่องรับบริการเชื่อมต่อไปยังเครื่องบริการ และร้องขอเอกสาร แล้วเครื่องบริการจัดส่งเอกสารที่ได้รับการร้องขอไปยังเครื่องรับบริการ และเครื่องรับบริการรับเอกสารนั้นมาแสดงผล โดยกรณีที่เครื่องบริการจัดส่งเอกสารไปให้เครื่องรับบริการไม่ได้ เช่น มีข้อขัดข้องทางเทคนิค เครื่องบริการจะส่งรายละเอียดเกี่ยวกับการขัดข้องเพื่อแจ้งให้ทราบด้วย หลังจากนั้น เครื่องรับบริการและเครื่องบริการจะเลิกการติดต่อ



ทั้งนี้ หากผู้ใช้คลิกเลือกการเชื่อมโยงใด ๆ ก็จะเข้าสู่กระบวนการดังกล่าวทั้งหมดใหม่ การเชื่อมต่อด้วยเอชทีทีพี ซึ่งยุติการติดต่อเมื่อเครื่องรับบริการได้รับเอกสารและแสดงผลเอกสารนี้ ช่วยให้ผู้อื่นสามารถเข้าถึงทรัพยากรต่าง ๆ บนเครื่องบริการได้ และเป็นการใช้การเชื่อมต่ออันจำกัดของเครื่องบริการได้อย่างสูงสุด

ไมม์ *(Multipurpose Internet Multimedia Extension - MIME)*

ไมม์ เป็นมาตรฐานซึ่งเป็นเกณฑ์ใช้ระบุประเภทข้อมูลในเวิลด์ไวด์เว็บ และในการประยุกต์ใช้ต่าง ๆ ของอินเทอร์เน็ต ในเบื้องต้น ไมม์ พัฒนาขึ้นเพื่อใช้บรรยายสารนิเทศที่ส่งทางไปรษณีย์อิเล็กทรอนิกส์ อย่างไรก็ดีมีการนำไมม์มาใช้กับเวิลด์ไวด์เว็บด้วย เพราะในการแสดงผลแฟ้มข้อมูลอย่างถูกต้อง เว็บเบราว์เซอร์จำเป็นต้องทราบว่าแฟ้มข้อมูลนั้นอยู่ในรูปแบบใด ซึ่งการตีความ

จะขึ้นอยู่กับประเภทข้อมูลของแฟ้มข้อมูลนั้น ไมม์จึงเป็นสิ่งที่ติดไปกับข้อมูลที่จัดส่ง เช่น แฟ้มข้อมูลที่มีนามสกุล .html จะมีประเภทของไมม์เป็น text / html เมื่อเบราว์เซอร์ได้รับแฟ้มข้อมูลประเภทดังกล่าว ก็จะสามารถจัดการกับแฟ้มข้อมูลนั้นได้อย่างเหมาะสม โดยการคืนรูป (render) ให้เป็นข้อความที่เขียนด้วยภาษาเอชทีเอ็มแอล (HyperText Markup Language - HTML) แล้วแสดงผลบนจอภาพ

ไมม์แยกออกเป็น 2 ส่วน คือ ส่วนทั่วไป (generic part) และส่วนเฉพาะ (specific part) เช่น text / html และ image / jpeg ในที่นี้ ส่วนทั่วไป ได้แก่ text และ image ซึ่งเป็นส่วนที่บอกให้ทราบประเภทข้อมูลโดยรวมว่าเป็น ข้อความ และภาพ ส่วน html และ jpeg เป็นส่วนเฉพาะที่ระบุว่า ประเภทข้อมูลที่เป็นข้อความนั้น ใช้คำสั่งภาษาเอชทีเอ็มแอล และประเภทข้อมูลที่เป็นภาพนั้น อยู่ในรูปแบบเจเพ็ก (Joint Photographic Experts Group - JPEG) สำหรับประเภทของไมม์นั้น มีอยู่มากมาย และได้มีการกำหนดชุดมาตรฐานขึ้นใช้ ทั้งนี้มีข้อสังเกตว่า โดยทั่วไป ระบบคอมพิวเตอร์จำนวนมาก จะใช้ชื่อแฟ้มข้อมูล เป็นวิธีการอย่างคร่าว ๆ ในการบันทึกประเภทข้อมูลด้วย เช่น แฟ้มข้อมูลเอชทีเอ็มแอล จะใช้นามสกุล .htm หรือ .html

Z39.50, ISO23950

Z39.50 เป็นโพรโทคอลสำหรับการค้นคืนสารนิเทศข้ามระบบห้องสมุดต่าง ๆ และเป็นมาตรฐานสากลที่ช่วยให้ผู้ใช้สามารถสืบค้นเครื่องคอมพิวเตอร์ที่อยู่ห่างไกลได้โดยใช้การต่อประสาน (interface) ที่คุ้นเคยของฐานข้อมูลบนเครื่องคอมพิวเตอร์ของตน

มาตรฐานนี้ เป็นโพรโทคอลที่ใช้กำหนดระเบียบปฏิบัติ (procedure) และโครงสร้างสำหรับการเสนอคำร้องขอในการสืบค้น และการตอบรับต่อคำร้องขอระหว่างระบบ จึงประกอบไปด้วยกระบวนการทางสารนิเทศ 2 ส่วน ได้แก่ ส่วนเริ่ม (origin) ซึ่งเป็นส่วนที่สนับสนุนการร้องขอของระบบที่อยู่ห่างไกล และส่วนเป้าหมาย (target) ซึ่งเป็นส่วนตอบรับที่มีฐานข้อมูล และแปลคำร้องขอให้เป็นตรรกะ (logic) ของระบบฐานข้อมูลที่เป็นเป้าหมายนั้น แล้วส่งคืนระเบียนหรือชุดผลลัพธ์กลับไป

แม้ว่า Z39.50 จะไม่ใช่โพรโทคอลที่โปรแกรมค้นหา (search engine) บนอินเทอร์เน็ตใช้เพราะโปรแกรมค้นหาใช้เอชทีทีพี (Noerr 2000: 123) แต่ Z39.50 ก็เป็นมาตรฐานที่มีความครอบคลุม (comprehensive) ความซับซ้อน และความสามารถมากกว่าการสืบค้นด้วยเอชทีทีพี

ISO10160 และ ISO10161

ISO10160 และ ISO10161 เป็นมาตรฐานด้านการยืมระหว่างห้องสมุด (InterLibrary Loan - ILL) ที่เป็นไปในทำนองเดียวกับมาตรฐาน Z39.63 นำมาใช้สำหรับกำหนดเกณฑ์เพื่อทำให้สามารถสื่อสารแบบอัตโนมัติในการยืมระหว่างห้องสมุดระบบต่าง ๆ และสนับสนุนการจัดการรายการ (transaction) เพื่อการยืมระหว่างห้องสมุด

ISO10160 เป็นส่วนที่เรียกว่า "นิยามของบริการ" (Service Definition) ในขณะที่ ISO10161-1 และ ISO10161-2 เป็นส่วนที่เรียกว่า "ข้อกำหนดของโพรโทคอล" (Protocol Specification) ทั้ง ISO10160 และ ISO10161-1 พิมพ์เป็นครั้งแรกในปี พ.ศ. 2536 (ค.ศ. 1993) และพิมพ์ครั้งที่ 2 ในปี พ.ศ. 2540 (ค.ศ. 1997) ส่วน ISO10161-2 พิมพ์ขึ้นในปี พ.ศ. 2540 (ค.ศ. 1997)

มาตรฐานด้านการเข้ารหัส

เนื่องจากคอมพิวเตอร์จัดเก็บอักขระในรูปการจัดเรียงลำดับบิต (sequence of bits) มาตรฐานในการเข้ารหัส จึงเป็นมาตรฐานที่เกี่ยวข้องกับวิธีการแทน (represent) อักขระ (character) แต่ละตัวในแฟ้มข้อมูลและระเบียน มาตรฐานในการเข้ารหัสที่ควรรู้จักมีดังต่อไปนี้

แอสกี *(American Standard Code for Information Interchange - ASCII, ISO646)*

แอสกี เป็นชุดอักขระที่มีการนำมาใช้อย่างกว้างขวาง เดิมที แอสกีจะแทนอักขระแต่ละตัวแบบ 7 บิต เช่น อักขระ A จะเข้ารหัสด้วยการจัดเรียงลำดับบิตแบบ 7 บิตได้เป็น 1000001 หรือหากพิจารณาเป็นเลขโดดฐานสอง คือ หมายเลข 65 นั่นเอง (Arms 2000: 169) อย่างไรก็ตามเนื่องด้วยการแทนรหัสแบบ 7 บิต จะแทนค่าได้เพียง 128 อักขระเท่านั้น ซึ่งในความเป็นจริงแล้ว จะใช้ได้เพียง 96 อักขระ คือ 32-127 ซึ่งเรียกว่าเป็น "ชุดรหัสแอสกีที่สามารถพิมพ์ได้" (printable ASCII character set) ส่วน 0-31 เป็นอักขระควบคุม (control character)

ต่อมา ได้มีการขยายชุดรหัสแอสกีเพิ่มเติม เป็นแบบ 8 บิต ทำให้มีจำนวนอักขระเพิ่มขึ้น รวมเป็น 256 อักขระ คือ 0-255 ชุดรหัสแบบ 8 บิตนี้ ครอบคลุมอักขระที่จำเป็นในภาษาอังกฤษ และมีอักขระพิเศษหลายตัวด้วย

ยูนิโค้ด *(Universal Character Set - Unicode, ISO10646)*

แม้ว่าจะมีชุดรหัสแอสกีเพื่อให้ใช้แทนอักขระได้ถึง 256 อักขระแล้ว แต่จำนวนอักขระก็ไม่เพียงพอต่อความต้องการใช้ในหลายภาษา เช่น ภาษากรีก เกาหลี จีน ญี่ปุ่น หรือแม้กระทั่งภาษาอังกฤษแบบเก่า (Old English) รวมถึงบางสาขาวิชา เช่น คณิตศาสตร์ เคมี และดนตรี ดังนั้น จึงมีความพยายามคิดค้นมาตรฐานใหม่เพื่อรองรับความต้องการนี้ แต่ก็ไม่มีมาตรฐานใด มีอักขระเพียงพอ จนกระทั่งมีมาตรฐานยูนิโค้ดเกิดขึ้น

ยูนิโค้ด เป็นเกณฑ์ในการเข้ารหัสสากลแบบ 16 บิต แทนค่าได้มากกว่า 65,000 อักขระ จึงเป็นมาตรฐานที่ให้หมายเลขเฉพาะสำหรับอักขระได้ทุกตัว ไม่ว่าจะเป็นแพลตฟอร์ม (platform) ใด โปรแกรมใด หรือภาษาใด รวมถึงเครื่องหมายต่าง ๆ เป็นจำนวนมาก ทั้งนี้ประมาณ 1 ใน 3 ของอักขระนั้น ยังไม่มีการกำหนด จึงยังมีรหัสว่างเพื่อการเพิ่มเติมอักขระใน

อนาคตได้อีก นอกจากนี้ ยูนิโค้ดยังเป็นมาตรฐานที่ใช้ง่าย สมบูรณ์ ทำให้ข้อมูลส่งผ่านระหว่างระบบต่าง ๆ ได้โดยไม่เกิดความเสียหาย และเป็นมาตรฐานที่มีการยอมรับนำมาใช้อย่างกว้างขวางโดยผู้นำอุตสาหกรรม เช่น Apple HP IBM Microsoft Oracle SAP หรือ Sun รวมทั้งมาตรฐานใหม่ต่าง ๆ เช่น Java JavaScript WML และ XML

ในการพัฒนามาตรฐานยูนิโค้ด มีการพิจารณาถึงความสัมพันธ์ระหว่างยูนิโค้ดกับซอฟต์แวร์ที่มีอยู่ด้วย เพื่อการยอมรับมาตรฐานดังกล่าว เพราะหากโปรแกรมคอมพิวเตอร์ทุกโปรแกรมต้องมีการเปลี่ยนแปลง คงไม่มีการยอมรับนำยูนิโค้ดมาใช้ ด้วยเหตุนี้ จึงมีการแทนอักขระของยูนิโค้ดแบบพิเศษเกิดขึ้น เรียกว่า "ยูทีเอฟ-8" (UTF-8)

ยูทีเอฟ-8 เป็นรูปแบบการเข้ารหัสที่มีประโยชน์ เพราะอักขระของยูนิโค้ดซึ่งสอดคล้องกับชุดรหัสแอสกี จะเข้ารหัสแบบ 1 ไบต์ มี 8 บิต เหมือนกัน ดังนั้น จึงสามารถตีความค่าไบต์ (byte value) ที่เหมือนกันนี้ เป็นอักขระยูนิโค้ดในรูปแบบยูทีเอฟ-8 หรือแอสกีก็ได้ เช่น ข้อความเอกสารเอชทีเอ็มแอล ที่สร้างขึ้นโดยใช้อักขระแอสกี จะไม่จำเป็นต้องมีการแก้ไขเพิ่มเติมใด ๆ ในการนำมาใช้กับโปรแกรมที่คาดหมายให้ใช้กับข้อมูลซึ่งเข้ารหัสแบบยูทีเอฟ-8 ในขณะเดียวกันอักขระของยูนิโค้ดแบบยูทีเอฟ-8 ก็สามารถใช้กับซอฟต์แวร์ต่าง ๆ ที่มีอยู่แล้วได้โดยไม่ต้องเขียนซอฟต์แวร์ใหม่เพิ่มเติม ยูทีเอฟ-8 เป็นรูปแบบที่นิยมใช้กับเอกสารเอชทีเอ็มแอลและโพรโทคอลอื่นในทำนองเดียวกัน นอกจากนี้ มาตรฐานยูนิโค้ด ยังมีรูปแบบการเข้ารหัสข้อมูลให้เป็นแบบยูทีเอฟ-16 (UTF-16) และยูทีเอฟ-32 (UTF-32) ด้วย

มอก.620 (TIS620)

มอก.620 เป็นมาตรฐานชุดอักขระภาษาไทยที่ใช้กันในราชอาณาจักรไทย กำหนดขึ้นโดยสำนักงานมาตรฐานผลิตภัณฑ์อุตสาหกรรมไทย (Thai Industrial Standards Institute - TISI) ในปี พ.ศ. 2529 (ค.ศ. 1986) และมีการปรับปรุงในปี พ.ศ. 2533 (ค.ศ. 1990) โดยเป็นมาตรฐานที่ได้รับการนำมาใช้อย่างกว้างขวางตั้งแต่นั้นเป็นต้นมา (Trin Tantsetthi 1998)

มอก.620 เป็นมาตรฐานที่พัฒนาขึ้น 2 รูปแบบ คือ แบบที่พัฒนาจากมาตรฐานแอสกีหรือ ISO646 โดยเพิ่มเติมอักขระภาษาไทยเข้าไปในส่วนหลังของการเข้ารหัสแบบ 8 บิต และแบบที่พัฒนาจากมาตรฐานไอบีเอ็ม เอบซีดิก (IBM EBCDIC) ซึ่งเอบซีดิก ย่อมาจาก Extended Binary Coded Decimal Interchange Code (EBCDIC) แต่เนื่องจากมาตรฐานแบบหลังไม่เป็นที่นิยมใช้นัก รูปแบบแรกจึงเป็นตัวแทนของมาตรฐานมอก.620ในที่สุด (ธวัชชัย ศรีสุเทพ 2544: 295-296)

มาตรฐานนี้ ได้รับการจดทะเบียนภายใต้ ISO2375 กับ European Computing Manufacturer's Association (ECMA) เป็นมาตรฐาน ISO-IR-166 รวมถึงการจดทะเบียนกับ Internet Assigned Numbers Authority (IANA) แล้ว นอกจากนี้ มอก.620 ยังเป็นมาตรฐานที่ผู้ค้าต่าง ๆ ได้นำไปพัฒนาเพิ่มเติมและมีการนำมาใช้ เช่น IBM Thai หรือ CP874 Mac Thai และ MS874 หรือ Windows-874

มาตรฐานด้านการแสดงรายละเอียดทรัพยากร

รูปแบบการแสดงรายละเอียดทรัพยากรที่ห้องสมุดดิจิทัลควรพิจารณา ซึ่งจะกล่าวถึงในที่นี้ แบ่งออกเป็น 2 ประเภท คือ การแสดงรายละเอียดโครงสร้าง (structure) ด้วยการใช้ Markup Language และการแสดงรายละเอียดโครงร่าง (layout) ด้วยการใช้ Page Description Language (PDL)

Markup Language

Markup Language เป็นคำที่มีที่มามาจากอุตสาหกรรมการพิมพ์ โดยบรรณาธิการจะเขียนรหัสหรือสัญลักษณ์ไว้ที่ขอบกระดาษของหนังสือหรือหนังสือพิมพ์ เพื่อระบุให้ผู้เรียงพิมพ์ทราบวิธีการที่จะแสดงลักษณะเฉพาะต่าง ๆ สำหรับในที่นี้ Markup Language หมายถึง มาตรฐานที่ใช้ในการแสดงรายละเอียดสารนิเทศ (เนื้อหา) และโครงสร้าง ประกอบไปด้วยแท็ก (tag) ต่าง ๆ โดยไม่ใช่ภาษาที่ใช้ในการแสดงรายละเอียดโครงร่าง

<u>เอสจีเอ็มแอล (Standard Generalized Markup Language - SGML, ISO8879)</u>

เอสจีเอ็มแอล เป็นภาษาที่กำหนดหลักเกณฑ์ทั่วไป และลักษณะโครงสร้างพื้นฐานของเอกสาร ซึ่งแยกการระบุเนื้อหาของเอกสารออกจากรูปแบบการแสดงผล ใช้ในการระบุส่วนต่าง ๆ เชิงโครงสร้าง เช่น หัวข้อ (heading) และข้อความเนื้อหา (body text) เป็นต้น พัฒนาการของภาษาเอสจีเอ็มแอล เริ่มขึ้นในปี พ.ศ. 2514 (ค.ศ. 1696) โดยทีมนักวิจัยของบริษัท IBM และเป็นมาตรฐานเปิดในปี พ.ศ. 2529 (ค.ศ. 1986)

ภาษาเอสจีเอ็มแอลมีส่วนประกอบแยกออกเป็น 3 ส่วน ได้แก่ ส่วนแรกเรียกว่า "Declaration" ใช้เพื่อแสดงรายละเอียดภาวะแวดล้อมในการประมวลผลที่จำเป็น และอาจมีสารนิเทศเกี่ยวกับชุดอักขระที่ใช้ ส่วนที่สอง เรียกว่า "Document Type Definition" (DTD) ซึ่งแสดงชุดแท็กสำหรับการแสดงรายละเอียดโครงสร้างและเนื้อหาของประเภทเอกสาร และส่วนสุดท้ายเป็นส่วนเอกสาร (document) ทั้งนี้ เนื่องจากเอกสารต่าง ๆ มีมากมายหลายประเภท การกำหนดชุดแท็กให้มีความครอบคลุมจึงเป็นไปได้ยาก มาตรฐานเอสจีเอ็มแอล จึงยอมให้ผู้ใช้กำหนดชุดแท็กสำหรับประเภทของเอกสารได้

เอชทีเอ็มแอล (HyperText Markup Language - HTML)

เอชทีเอ็มแอล เป็นภาษาที่มีจุดเริ่มต้นในปี พ.ศ. 2533 (ค.ศ. 1990) โดยมีพื้นฐานมาจากภาษาเอสจีเอ็มแอล เนื่องจากเอสจีเอ็มแอลเป็นภาษาที่ซับซ้อน เอชทีเอ็มแอลจึงเป็นภาษาที่ง่ายกว่า อย่างไรก็ดี แม้ว่าในเบื้องต้น ภาษาเอชทีเอ็มแอล จะเป็นภาษาที่มีความสัมพันธ์ในเชิงโครงสร้าง แต่ต่อมา เอชทีเอ็มแอล ก็ได้กลายเป็นภาษาที่มีการผสมผสานสารนิเทศเชิงรูปแบบการแสดงผลเข้ากับสารนิเทศเชิงโครงสร้างเมื่อมีการเพิ่มคุณลักษณะต่าง ๆ ที่ช่วยให้นักออกแบบเว็บเพจสามารถควบคุมวิธีการแสดงผลได้ จึงอาจกล่าวได้ว่า เอชทีเอ็มแอล เป็นภาษาที่ใช้ในการเขียนเว็บเพจ เพื่อกำหนดโครงสร้างและรูปแบบการแสดงผลของเอกสาร รวมทั้งการเชื่อมโยง

แฟ้มข้อมูลเอชทีเอ็มแอลมีนามสกุล .htm หรือ .html โดยแท็กในภาษาเอชทีเอ็มแอลส่วนใหญ่จะปรากฏเป็นคู่ คือ แท็กเปิด < > และ แท็กปิด </ > ตัวอย่างของแท็กที่ใช้กำหนดโครงสร้าง เช่น ... ซึ่งระบุย่อหน้า และ<h1>...</h1> ซึ่งระบุหัวข้อระดับที่ 1 และตัวอย่างของแท็กที่ใช้กำหนดรูปแบบการแสดงผล เช่น ... ซึ่งระบุให้อักษรหรือข้อความเป็นตัวหนา และ ... ซึ่งระบุให้อักษรหรือข้อความเป็นตัวเอียง ส่วนแท็กที่ใช้กำหนดการเชื่อมโยงซึ่งรู้จักกันดี คือ <a>... ทั้งนี้ แท็กของเอชทีเอ็มแอล อาจเป็นแท็กเดี่ยวที่ไม่ปรากฏเป็นคู่ก็ได้ เช่น
 ซึ่งระบุให้ขึ้นบรรทัดใหม่ และ<hr> ซึ่งระบุให้ใส่เส้นคั่น

ทั้งนี้ เนื่องจากแท็กในภาษาเอชทีเอ็มแอล เป็นแท็กที่มีการกำหนดไว้ก่อน (predefine) มีความหมายเฉพาะ จึงไม่มีความยืดหยุ่น เอชทีเอ็มแอลจึงเป็นภาษาที่มีข้อจำกัด นอกจากนี้ เอชทีเอ็มแอลยังมีข้อจำกัดในการสร้างเว็บเพจที่สามารถโต้ตอบกับผู้ใช้ การเชื่อมโยงข้อมูลบนฐานข้อมูล ซึ่งสามารถแสดงข้อมูลข่าวสารใหม่ ๆ ได้อย่างทันท่วงที และการคำนวณประมวลผล จึงมีการนำเทคโนโลยีอื่นมาใช้ประกอบเอชทีเอ็มแอล เช่น Java JavaScript และ Perl เป็นต้น

เอ็กซเอ็มแอล (eXtensible Markup Language - XML)

เอ็กซเอ็มแอล เป็นมาตรฐานที่คิดค้นขึ้นในปี พ.ศ. 2539 (ค.ศ. 1996) โดยมีพัฒนาการมาจากภาษาเอสจีเอ็มแอล มุ่งหมายให้ใช้ในการระบุส่วนที่เป็นเนื้อหามากกว่าส่วนโครงร่างหรือการแสดงผล เอ็กซเอ็มแอล จึงไม่มีคำสั่งด้านการจัดรูปแบบภายในตัวเอง และใช้มาตรฐานเอ็กซเอสแอล (eXtensible Stylesheet Language - XSL) ในการกำหนดโครงร่างสำหรับชุดอักขระที่เอ็กซเอ็มแอลใช้นั้น คือ มาตรฐานยูนิโค้ด ทั้งนี้ กฎระเบียบของแฟ้มข้อมูลเอ็กซเอ็มแอล ซึ่งมีนามสกุล .xml มีความเข้มงวด ดังจะเห็นได้ว่า ในเอกสารเอ็กซเอ็มแอล

แท็กทั้งหมดต้องเป็นอักษรตัวเล็ก (lowercase) แท็กเปิดทั้งหมดต้องมีแท็กปิด การซ้อนกัน (nest) ของแท็ก ต้องเป็นไปอย่างถูกต้องเหมาะสม (Tennant 2001: 30) เป็นต้น

อย่างไรก็ดี เอ็กซเอ็มแอลก็เป็นภาษาที่ง่ายกว่าภาษาเอสจีเอ็มแอลซึ่งมีความซับซ้อน ในขณะที่มีความยืดหยุ่นมากกว่าภาษาเอชทีเอ็มแอลด้วย เพราะเอ็กซเอ็มแอลได้รับการคิดค้นขึ้นมาเพื่อให้อนุวัต (implement) ได้ง่ายเหมือนกับภาษาเอชทีเอ็มแอล แต่ก็มีความสามารถเพียงพอต่อการประยุกต์ใช้ขั้นสูง (Baker 1999) นอกจากนี้ เอ็กซเอ็มแอล ยังมีลักษณะขยายได้ (extensible) เพราะไม่มีชุดแท็กที่กำหนดไว้ก่อน หากแต่เป็นเพียงเค้าโครง (framework) ซึ่งสามารถเพิ่มเติมได้ (Deegan and Tanner 2002: 126) เอ็กซเอ็มแอลมีความยืดหยุ่น เพราะในขณะที่เอชทีเอ็มแอลกำหนดว่าแท็กและแอ็ททริบิวต์ (attribute) หมายถึงอะไร รวมทั้งข้อความระหว่างแท็กนั้น จะแสดงผลอย่างไร เอ็กซเอ็มแอลจะใช้แท็กเพื่อกำหนดชิ้นส่วนของข้อมูล และปล่อยให้การตีความข้อมูลนั้นเป็นไปตามการประยุกต์ใช้ (application) ที่อ่านข้อมูลนั้น กล่าวคือ หากเห็น <p> ในแฟ้มข้อมูลเอ็กซเอ็มแอล <p> อาจมิได้หมายถึงย่อหน้า แต่อาจเป็นราคา (price) พารามิเตอร์ (parameter) หรือบุคคล (person) เป็นต้น ขึ้นอยู่กับบริบท (W3C Communications Team 2001)

นอกจากการอนุวัตได้ง่ายแล้ว เอ็กซเอ็มแอลยังได้รับการออกแบบมาเพื่อให้สามารถทำงานระหว่างกันกับทั้งภาษาเอสจีเอ็มแอล และภาษาเอชทีเอ็มแอล รวมถึงมีความสามารถในการย้ายจากภาษาเอชทีเอ็มแอลมายังภาษาเอ็กซเอ็มแอลได้ โดยไม่ก่อให้เกิดความเสียหาย ทั้งนี้ เนื่องจากมีวัสดุต่าง ๆ ที่เขียนขึ้นด้วยภาษาเอชทีเอ็มแอลเป็นจำนวนมาก หากมาตรฐานใหม่เข้ากันไม่ได้กับเอชทีเอ็มแอล จึงอาจไม่เป็นที่ยอมรับ อนึ่ง ในปี พ.ศ. 2543 (ค.ศ. 2000) World Wide Web Consortium (W3C) ได้ประกาศใช้มาตรฐานเอ็กซเอชทีเอ็มแอล (eXtensible HyperText Markup Language - XHTML) ซึ่งเป็นมาตรฐานที่ผสมผสานระหว่างเอ็กซเอ็มแอลและเอชทีเอ็มแอลขึ้นใช้ด้วย

อีเอดี (Encoded Archival Description - EAD)

อีเอดี เป็นมาตรฐานที่มีจุดเริ่มต้น ณ University of California, Berkeley ในปี พ.ศ. 2536 (ค.ศ. 1993) และมีการประกาศใช้ Version 1.0 ในปี พ.ศ. 2541 (ค.ศ. 1998) อีเอดี เป็นมาตรฐานที่มีพื้นฐานมาจากภาษาเอสจีเอ็มแอล ใช้เป็นมาตรฐานสำหรับเครื่องมือช่วยค้นจดหมายเหตุ (archival finding aids) โดยเครื่องมือช่วยค้น หมายถึง บัญชีรายการ (inventory) ทะเบียน (register) ดรรชนี (index) หรือคู่มือ (guide) ที่จัดทำขึ้นโดยแหล่งรวบรวมจดหมายเหตุและหนังสือตัวเขียน (archival and manuscript repository) เพื่อให้สารนิเทศ

อย่างละเอียดเกี่ยวกับทรัพยากรสารนิเทศใดโดยเฉพาะ (EAD 2000) อันเป็นการสนับสนุนหรือช่วยในการใช้ทรัพยากรที่มีอยู่ (holding) ได้

ทั้งนี้ เนื่องด้วยจดหมายเหตุเป็นการรวบรวมชุดเอกสารที่มีความพิเศษ (unique) และเป็นการบันทึกของบุคคล ครอบครัว หรือองค์กร การแสดงรายละเอียดทรัพยากรสารนิเทศที่เป็นจดหมายเหตุ จึงมีความซับซ้อน มีลักษณะเป็นลำดับชั้น (hierarchical) มีความยาว และเป็นเชิงวิเคราะห์ (analytical) ต่างจากการแสดงรายละเอียดทางบรรณานุกรม (bibliographic description) การใช้ภาษาเอสจีเอ็มแอล และต่อมา คือ ภาษาเอ็กซเอ็มแอล จึงมีความเหมาะสมมากกว่า MARC (Machine-Readable Cataloging)

ปัจจุบัน อีเอดี เป็นมาตรฐานที่ดูแลรักษาโดย Network Development and MARC Standards Office ของ Library of Congress (LC) ร่วมกันกับ Society of American Archivists และเป็นมาตรฐานที่ยอมรับกันอย่างกว้างขวางในวงการจดหมายเหตุ ตัวอย่างของโครงการที่ใช้อีเอดีนั้น เช่น The Making of America II Project และ Bodlein Library in Oxford เป็นต้น

ทีอีไอ (Text Encoding Initiative - TEI)

ทีอีไอ เป็น DTD ที่ใช้ภาษาเอสจีเอ็มแอลเช่นเดียวกับอีเอดี ถือกำเนิดขึ้นในปี พ.ศ. 2530 (ค.ศ. 1987) เพื่อพัฒนาแนวทางสำหรับการเข้ารหัสข้อความที่เครื่องอ่านได้ ซึ่งเป็นที่สนใจในสาขามนุษยศาสตร์และสังคมศาสตร์ (The TEI FAQ 2001) มาตรฐานนี้ เป็นมาตรฐานระดับสากลและสหสาขาวิชาที่ช่วยห้องสมุด พิพิธภัณฑ์ สำนักพิมพ์ และนักวิชาการในการแทนข้อความทางวรรณกรรมและภาษา (literary and linguistic text) ทุกประเภท เพื่อการวิจัยและการสอนแบบออนไลน์ (online research and teaching) โดยใช้เกณฑ์ในการเข้ารหัสที่สื่อความหมายได้มากที่สุด และล้าสมัยได้น้อยที่สุด (Welcome to the TEI Website 2002) ทั้งนี้ ทีอีไอ ให้เกณฑ์ในการเข้ารหัสข้อความเพื่อที่จะสามารถกำหนดส่วนต่าง ๆ เช่น จุดเริ่มต้นและจุดจบของบรรทัด ย่อหน้า หน้าเอกสาร บท หรือองก์ โดยที่ข้อความดังกล่าว จะสามารถนำมาประมวลผล เพื่อสร้างดรรชนีที่ถูกต้องสำหรับการสืบค้นได้

มาตรฐานทีอีไอฉบับ "P3" ได้รับการเผยแพร่ในปี พ.ศ. 2537 (ค.ศ. 1994) และกลายเป็นมาตรฐานแบบพฤตินัยสำหรับการเข้ารหัสข้อความทางวรรณกรรมและภาษา รวมถึงสิ่งต่าง ๆ ในทำนองเดียวกันนี้ในที่สุด สำหรับมาตรฐานฉบับปัจจุบันของทีอีไอ มีชื่อเรียกว่า "P4" วัตถุประสงค์หลักของฉบับปรับปรุงแก้ไขนี้ คือ เพื่ออนุวัตการสนับสนุนภาษาเอ็กซเอ็มแอลที่เหมาะสม โดยยังคงให้ความมั่นใจได้ว่า เอกสารที่จัดทำขึ้นด้วยข้อกำหนดของทีอีไอฉบับก่อนหน้าจะยังคงใช้ได้กับฉบับใหม่นี้ (The TEI Guidelines 2002) นอกจากนี้ ทีอีไอ ยังมีมาตรฐานที่เรียกว่า

"TEI Lite" อันประกอบด้วยชุดย่อย (subset) ของชุดแท็กทีอีไอทั้งหมด ซึ่งเลือกให้มีแท็กที่จำเป็นโดยทั่วไปที่สุด

ทีอีไอ ได้รับความสนับสนุนโดย Association for Computers and the Humanities (ACH) Association for Computational Linguistics (ACL) และ Association for Literary and Linguistic Computing (ALLC) และเป็นมาตรฐานที่มีโครงการต่าง ๆ มากมายนำไปใช้ เช่น American Memory form the Library of Congress British National Corpus Cambridge University Press และ Digital Dictionary of Budhhism เป็นต้น

Page Description Language (PDL)

Page Description Language เป็นมาตรฐานที่ใช้ในการแสดงรายละเอียดรูปร่างของหน้าเอกสาร หรือโครงร่าง ไม่ใช่เนื้อหาหรือโครงสร้าง โดยมีความมุ่งหมายเพื่อนำเสนอภาพ (image) หน้าเอกสารที่สมบูรณ์ ได้รับการจัดระเบียบแล้ว และเป็นรูปแบบสุดท้าย บนอุปกรณ์การพิมพ์หรืออุปกรณ์การแสดงผลอื่น ๆ โดยมีคุณภาพและความถูกต้องเชิงกราฟิกส์ (graphic quality and precision) เหมือนกับเอกสารที่จัดพิมพ์เป็นอย่างดีด้วยวิธีการดั้งเดิม

<u>โพสต์สคริปต์ (PostScript)</u>

โพสต์สคริปต์ เป็นภาษาการเขียนโปรแกรม (programming language) ที่พัฒนาขึ้นโดยบริษัท Adobe Systems ใช้ในการควบคุมการพิมพ์ เพื่อคืนรูปลักษณะที่ถูกต้องของเอกสารสำหรับการพิมพ์บนแพลตฟอร์มใด ๆ ทั้งนี้ มาตรฐานโพสต์สคริปต์ มีข้อดีในการให้งานพิมพ์สี (color output) ที่มีความละเอียดสูง อย่างไรก็ดี แฟ้มข้อมูลเชิงกราฟิกส์ของโพสต์สคริปต์นั้นค่อนข้างยากและไม่เป็นไปในเชิงปฏิบัติ (impractical) ต่อการแก้ไขหรือปรับเปลี่ยน (Dempsey and others 1998)

ปัจจุบัน โพสต์สคริปต์ได้พัฒนาจาก PostScript Level 2 มาเป็น PostScript 3 ซึ่งแม้ว่าจะไม่มีการเปลี่ยนแปลงที่สำคัญ แต่มีข้อได้เปรียบ คือ สนับสนุนการไล่โทนสีเทาได้ถึง 4,096 ระดับสี ในขณะที่อดีต จะมีการจำกัดระดับสีเทาไว้ไม่เกิน 256 ระดับสีเท่านั้น มีการเพิ่มจำนวนแบบอักษรหลักเป็น 136 แบบ สนับสนุนการใช้แฟ้มข้อมูลพีดีเอฟ (Portable Document Format - PDF) ที่สร้างจากโปรแกรม Acrobat ของ Adobe ได้โดยตรง เพิ่มความสามารถด้านกราฟิกส์ ได้แก่ การผสมภาพบริเวณรอยต่อให้ดูกลมกลืน และการรองรับภาพแบบ 3 มิติ นอกจากนี้ ในการพิมพ์ออกเวิลด์ไวด์เว็บ ได้มีการเพิ่มฟังก์ชัน (function) ในการใช้งานอินเทอร์เน็ตเข้าไปในโพสต์สคริปต์ด้วย (ทรงลักษณ์ สกุลวิจิตร์สินธุ 2544)

พีดีเอฟ (Portable Document Format - PDF)

พีดีเอฟ เป็นพัฒนาการของภาษาการเขียนโปรแกรมโพสต์สคริปต์ ซึ่งคิดค้นขึ้นโดยบริษัท Adobe Systems เช่นเดียวกัน แฟ้มข้อมูลนั้น ใช้นามสกุล .pdf มาตรฐานนี้ใช้เป็นรูปแบบการแทนเอกสารที่ไม่ขึ้นกับการประยุกต์ใช้และระบบคอมพิวเตอร์ใด ๆ และเป็นรูปแบบแฟ้มข้อมูลสากลที่ยังคง (preserve) แบบอักษร การจัดรูปแบบ สี และภาพกราฟิกส์ ทั้งยังสนับสนุนการเชื่อมโยงทั้งภายในและระหว่างเอกสาร บรรณนิทัศน์ (annotation) โครงบท (chapter outline) การสืบค้น และคุณลักษณะอื่น ๆ ที่เป็นไปได้ทางอิเล็กทรอนิกส์ รวมถึงการพิมพ์ ซึ่งมีวิธีการที่ผู้สร้างเอกสารสามารถป้องกันผู้ใช้จากการพิมพ์หรือการใช้เอกสารไปในทางที่ผู้สร้างไม่อนุญาตได้ด้วย

พีดีเอฟ เป็นมาตรฐานที่มีการนำมาใช้อย่างกว้างขวางบนเวิลด์ไวด์เว็บ โดยการสร้าง แก้ไข และรับชมเอกสารพีดีเอฟ จะดำเนินการได้ด้วยการใช้ชุดซอฟต์แวร์ที่เรียกว่า "Acrobat" เช่น Acrobat Reader (ผู้ใช้สามารถดาวน์โหลด (download) ได้โดยไม่เสียค่าใช้จ่าย) Acrobat PDF Writer Acrobat Exchange Acrobat Distiller และ Acrobat Search เป็นต้น

มาตรฐานด้านสื่อประเภทต่าง ๆ

สื่อที่ใช้บนเวิลด์ไวด์เว็บมีมากมายหลายประเภท เพราะเวิลด์ไวด์เว็บสามารถนำเสนอสารนิเทศได้หลากรูปแบบ ทั้งนี้ ภายใต้สารนิเทศแต่ละรูปแบบ ยังประกอบด้วยมาตรฐานต่าง ๆ อีกเป็นจำนวนมาก ในการดำเนินงานห้องสมุดดิจิทัล ความรู้ความเข้าใจในมาตรฐานเหล่านี้จึงเป็นสิ่งจำเป็น เพื่อการเลือกใช้มาตรฐานได้อย่างเหมาะสม

สื่อภาพกราฟิกส์ (graphic image)

นอกจากมาตรฐาน Tagged Image File Format (TIFF) ซึ่งเป็นรูปแบบของภาพที่มักใช้กับเครื่องกราดภาพ (scanner) แล้ว มาตรฐานสื่อภาพกราฟิกส์ที่ควรรู้จักสำหรับห้องสมุดดิจิทัล คือ จิฟ (Graphics Interchange Format - GIF) เจเพ็ก (Joint Photographic Experts Group - JPEG) และปิง (Portable Network Graphics - PNG)

จิฟ (Graphics Interchange Format - GIF)

จิฟเป็นมาตรฐานภาพกราฟิกส์ ที่พัฒนาโดยบริษัท CompuServe Information Service ไม่ขึ้นกับระบบปฏิบัติการใด ๆ แสดงผลได้ 256 สี (8 บิต) และเว็บเบราว์เซอร์ส่วนใหญ่สนับสนุน จิฟเหมาะกับภาพที่มีสีพื้น ๆ ไม่ซับซ้อน เป็นภาพเชิงเส้น เช่น โลโก้ (logo)

การ์ตูน แผนภูมิ และปุ่ม ภาพที่ต้องการให้มีบางส่วนเป็นลักษณะโปร่งใส รวมถึงภาพที่เคลื่อนไหวได้แบบง่าย ๆ

แฟ้มข้อมูลจิฟ แบ่งออกได้เป็น 2 ประเภท คือ GIF 87a ซึ่งใช้การบีบอัด (compression) แบบ Lempel Zev Welch หรือ LZW อันเป็นระบบการบีบอัดที่ไม่สูญเสีย (lossless) กล่าวคือ ไม่มีการสูญเสียข้อมูลของภาพไปในกระบวนการบีบอัด ทำให้หลังจากนำภาพมายกเลิกการบีบอัดแล้ว ภาพจะยังคงมีคุณภาพเช่นเดิม อย่างไรก็ดี การบีบอัดดังกล่าวเหมาะกับภาพที่มีสีไม่ซับซ้อน หรือมีจำนวนสีไม่มาก นอกจากนี้ GIF 87a ยังมีคุณสมบัติในการทำ interlacing ด้วย สำหรับแฟ้มข้อมูลจิฟอีกประเภท คือ GIF 89a มีลักษณะเหมือนกับ GIF 87a แต่มีคุณสมบัติเพิ่มเติมในการทำภาพโปร่งใส (transparency) และภาพเคลื่อนไหว (animation) แฟ้มข้อมูลทั้งสองประเภท ใช้นามสกุล .gif เหมือนกัน

โดยปกติ ภาพจิฟ ซึ่งเรียกว่า "Non-Interlaced GIF" หรือ "Standard GIF" จะดาวน์โหลดและแสดงผลทีละบรรทัดจากบนลงล่าง จนได้ภาพที่สมบูรณ์ อย่างไรก็ดี หากเป็นภาพจิฟแบบที่เรียกว่า "Interlaced GIF" นั้น จะดาวน์โหลดและแสดงผลภาพทั้งหมดในครั้งเดียวด้วยความละเอียดของภาพที่ต่ำ แล้วจึงค่อย ๆ มีความชัดเจนขึ้นเรื่อย ๆ จนได้ภาพที่สมบูรณ์ในที่สุด กล่าวคือ การดาวน์โหลดและแสดงผลภาพแบบนี้ จะแบ่งออกได้เป็น 4 ระดับ เริ่มจากการแสดงผลภาพหลังจากที่ดาวน์โหลดมาได้เพียง 12.5% หลังจากนั้น บริเวณที่ไม่ชัดเจนจะค่อย ๆ ชัดเจนขึ้นเมื่อมีการดาวน์โหลดข้อมูลเพิ่มขึ้นเป็น 25% 50% และ 100% (Niederst 1999: 268) ภาพจิฟแบบ interlaced GIF มีประโยชน์ในการช่วยให้ผู้ใช้ทราบขอบเขตของภาพที่กำลังดาวน์โหลดมา อันสามารถนำมาใช้ในการตัดสินใจต่าง ๆ ได้ เช่น การตัดสินใจว่าจะรอชมภาพที่กำลังดาวน์โหลดมาหรือไม่ และการตัดสินใจเลือกการเชื่อมโยงบนภาพที่เรียกว่า "Imagemap" เพื่อการเชื่อมโยงไปยังเว็บเพจอื่น ๆ ต่อไป โดยไม่จำเป็นต้องรอให้ดาวน์โหลดและแสดงผลภาพจนครบทั้งหมดเสียก่อน เป็นต้น อย่างไรก็ตาม ภาพจิฟแบบนี้ ควรใช้กับภาพที่มีขนาดใหญ่ ไม่ใช่ภาพขนาดเล็ก เพราะภาพขนาดเล็ก ใช้เวลาในการดาวน์โหลดและแสดงผลไม่นานอยู่แล้ว

นอกจากนี้ จิฟยังสามารถใช้สร้างภาพแบบที่เรียกว่า "Transparent GIF" ซึ่งเป็นภาพที่มีบางส่วนโปร่งใส ทำให้สามารถมองทะลุเห็นด้านหลังได้ เช่น สีหรือลวดลายของพื้นหลัง (background) และทำให้รูปทรงของภาพมีลักษณะอื่นนอกจากรูปทรงสี่เหลี่ยมได้ รวมทั้งทำให้เกิดความกลมกลืนมากยิ่งขึ้น ส่วนภาพจิฟที่เป็นภาพต่อเนื่องหรือภาพเคลื่อนไหวเรียกว่า "Animated GIF" โดยเป็นรูปแบบของแฟ้มข้อมูลภาพ ที่มีภาพหลายภาพ เป็นเฟรม (frame) ต่าง ๆ มาประกอบเข้าเป็นภาพเดียวกัน เมื่อแสดงผล ภาพจะเปลี่ยนไปตามเฟรมต่าง ๆ ในลักษณะต่อเนื่อง จึงทำให้รู้สึกว่า เป็นภาพที่มีความต่อเนื่องหรือเคลื่อนไหวได้ ภาพจิฟที่

เคลื่อนไหวได้นี้มีข้อดี คือ ผู้ใช้สามารถชมภาพแบบนี้ได้ด้วยเว็บเบราว์เซอร์ส่วนใหญ่ที่ใช้กันอยู่ในปัจจุบัน โดยไม่จำเป็นต้องใช้ซอฟต์แวร์พิเศษหรือตัวเสริม (plug-in) แต่อย่างใด ในด้านการสร้างภาพแบบนี้ ก็สามารถสร้างได้อย่างง่ายดายด้วยโปรแกรมต่าง ๆ เป็นจำนวนมาก อย่างไรก็ดี แม้ว่าภาพจิฟที่เคลื่อนไหวได้ จะสามารถสร้างความดึงดูดใจ และช่วยให้เกิดความมีชีวิตชีวา แต่ก็ควรพิจารณาเลือกใช้อย่างเหมาะสม ตามความจำเป็นด้วย มิเช่นนั้นอาจกลายเป็นการรบกวนสายตาของผู้ใช้ สร้างความสับสน ก่อให้เกิดความสนใจไปกับสิ่งที่ไม่สำคัญ และใช้เวลาในการดาวน์โหลดนานได้

เจเพ็ก (Joint Photographic Experts Group - JPEG, ISO10918-4)

เจเพ็กเป็นมาตรฐานภาพกราฟิกส์ที่ไม่ขึ้นกับระบบปฏิบัติการใด ๆ และเว็บเบราว์เซอร์ส่วนใหญ่สนับสนุนเช่นเดียวกับจิฟ แฟ้มข้อมูลมีนามสกุล .jpg หรือ .jpeg แสดงผลได้ 16.7 ล้านสี (24 บิต) หรือที่เรียกว่า "true color" ซึ่งใช้หลักการผสมสีของแสงสีแดง (Red) เขียว (Green) และน้ำเงิน (Blue) บนจอภาพ เรียกอย่างย่อว่า "อาร์จีบี" (RGB) แต่ละสีมีค่า 8 บิต รวม 3 สี จึงเป็น 24 บิต เจเพ็กเหมาะกับภาพถ่าย ภาพจากการกราดภาพ ภาพวาด ภาพสีน้ำ ภาพสเกลสีเทา (grayscale) หลายระดับ และภาพที่มีสีซับซ้อน

เจเพ็กใช้การบีบอัดภาพซึ่งเป็นเทคนิคเชิงคณิตศาสตร์ขั้นสูง เรียกว่า "Discrete Cosine Transformation" สามารถเลือกระดับของการบีบอัดที่จะใช้กับภาพได้ตามความต้องการ โดยบีบอัดให้ภาพเล็กลงกว่าแฟ้มข้อมูลต้นฉบับได้ถึง 100 เท่า อย่างไรก็ดี ควรพิจารณาด้วยว่า การบีบอัดภาพมาก ๆ นั้น จะทำให้ภาพมีคุณภาพที่ด้อยลงไปด้วย (Lynch and Horton 1999: 120) ในการบีบอัดภาพเจเพ็ก จึงควรหาขนาดของแฟ้มข้อมูลที่เล็กที่สุดซึ่งยังคงคุณภาพของภาพในระดับที่ยอมรับได้ ทั้งนี้ เนื่องจากการบีบอัดภาพเจเพ็ก เป็นระบบการบีบอัดที่สูญเสีย (lossy) กล่าวคือ มีการสูญเสียข้อมูลไปบางส่วนในกระบวนการบีบอัด รวมทั้งการบีบอัดและบันทึกภาพเจเพ็กจะเป็นการลดคุณภาพของภาพอย่างที่ไม่สามารถจะทำให้มีคุณภาพที่สมบูรณ์ดังเดิมได้ อีกทั้ง ทุกครั้งที่มีการปรับแต่งภาพแล้วบันทึกใหม่ ก็สามารถทำให้สูญเสียข้อมูล ตลอดจนการตกแต่งใด ๆ ยังอาจส่งผลให้ประสิทธิภาพในการบีบอัดครั้งถัดไปลดลง โดยแฟ้มข้อมูลมีขนาดใหญ่ขึ้น ในการบีบอัดภาพเจเพ็ก จึงควรเก็บแฟ้มข้อมูลภาพต้นฉบับที่ยังไม่ได้บีบอัดไว้ด้วยเสมอ

ปิง (Portable Network Graphics - PNG)

ปิงเป็นมาตรฐานที่ได้รับการพัฒนาขึ้นในปี พ.ศ. 2538 (ค.ศ. 1995) เพื่อมาแทนที่จิฟ เหมาะกับภาพที่โดยทั่วไปจัดเก็บด้วยมาตรฐานจิฟ โดยปิงใช้การบีบอัดที่เรียกว่า "Deflate" ซึ่งเป็นระบบการบีบอัดที่ไม่สูญเสีย พัฒนามาจากระบบ LZW ที่ใช้กับภาพจิฟ แต่มีการบีบอัดที่ดีกว่า แฟ้มข้อมูลจึงมีขนาดเล็กกว่าได้ นอกจากนี้ การบันทึกภาพปิงแล้วบันทึกใหม่อีก

จะไม่ทำให้คุณภาพของภาพลดลง จึงเป็นรูปแบบที่มีประโยชน์สำหรับการจัดเก็บในขั้นตอนระหว่างการแก้ไข (Dempsey and others 1998) ทั้งนี้ ปิงยังเหมาะที่จะนำมาใช้กับภาพที่ต้องการใช้คุณลักษณะเด่นที่น่าสนใจต่าง ๆ เช่น (Niederst 1999: 291, 293)

- การสนับสนุนการแสดงผลสีต่าง ๆ คือ 256 สี (8 บิต) เหมือนจิฟ ภาพสเกลสีเทา 65,536 ระดับ (16 บิต) ซึ่งทำให้สามารถจัดเก็บภาพขาวดำได้อย่างละเอียด และภาพที่แสดงผลแบบ "true color" ทั้งแบบ 16.7 ล้านสี (24 บิต) และกว่า 278 ล้านสี (48 บิต) โดยกรณี 48 บิต นั้น จะเป็นการผสมสีของแสงสีแดง เขียว และน้ำเงิน ซึ่งแต่ละสีมีค่า 16 บิต
- การทำให้ภาพโปร่งใสได้หลายระดับ คือ โปร่งใสแบบ 256 ระดับ (8 บิต) หรือ 65,536 ระดับ (16 บิต) ไม่ใช่แค่เพียงโปร่งใส หรือไม่โปร่งใสแบบจิฟ
- ความสามารถในการจัดเก็บข้อความ ซึ่งมีประโยชน์ในการทำให้ข้อความนั้นติดไปกับภาพอย่างถาวร เช่น รายละเอียดเกี่ยวกับภาพหรือสารนิเทศที่สำคัญดังเช่น ลิขสิทธิ์
- การดาวน์โหลดและแสดงผลภาพในลักษณะแบบ interlaced GIF ได้ โดยสามารถเติมเต็มสารนิเทศทั้งในด้านแนวนอน (horizontal) และแนวตั้ง (vertical) ในขณะที่ interlaced GIF จะดาวน์โหลดและแสดงผลด้วยการเติมเต็มสารนิเทศเฉพาะในด้านแนวนอนเท่านั้น

แม้ว่าปิงจะเป็นมาตรฐานสำหรับภาพกราฟิกส์ที่มีข้อดีต่าง ๆ มากมาย แต่ก็มีข้อเสียคือ เบราว์เซอร์ที่ใช้ในปัจจุบัน ยังไม่สนับสนุนคุณลักษณะต่าง ๆ ของปิงได้อย่างสมบูรณ์ จึงส่งผลกระทบต่อการแสดงผลได้ อย่างไรก็ดี สถานการณ์ดังกล่าวนี้ ก็มีแนวโน้มว่าจะดีขึ้น

สื่อเสียง (audio) และวีดิทัศน์ (video)

สื่อเสียงและวีดิทัศน์เป็นสื่อที่ปัจจุบันมีการนำมาใช้ประโยชน์บนเวิลด์ไวด์เว็บอย่างค่อนข้างแพร่หลาย แต่เนื่องด้วยขนาดของแฟ้มข้อมูลที่ทำให้การดาวน์โหลดนั้น มักใช้เวลานาน การใช้ทั้งสื่อเสียงและวีดิทัศน์ จึงต้องมีการพิจารณา โดยไม่ควรใช้มากจนเกินไป

แฟ้มข้อมูลเสียงที่ใช้กันอยู่บนเวิลด์ไวด์เว็บ สามารถแบ่งออกได้เป็น 2 ประเภท เรียกว่า "Non-Streaming Audio" หรือ "Static Audio" และ "Streaming Audio" โดยแฟ้มข้อมูลประเภทแรกนั้น จะต้องดาวน์โหลดมายังฮาร์ดดิสก์ (hard disk) ของผู้ใช้ทั้งหมดเสียก่อน จึงจะเริ่มเล่นได้ ส่งผลให้อาจต้องใช้เวลาในการดาวน์โหลดนาน สำหรับแฟ้มข้อมูลประเภทหลัง จะเริ่มเล่นแทบจะทันทีที่มีการร้องขอ และยังคงเล่นต่อไปในระหว่างที่มีการส่งข้อมูล จึงเป็นการช่วยลดปัญหาเรื่องเวลาที่ใช้ในการดาวน์โหลดได้ และเหมาะกับแฟ้มข้อมูลขนาดใหญ่ อย่างไรก็ดี แฟ้มข้อมูล

Streaming Audio มีข้อเสียคือ การสร้างแฟ้มข้อมูลประเภทนี้มีความซับซ้อนกว่าและซอฟต์แวร์ที่ใช้ในการสร้างนั้นก็มีราคาแพงกว่าด้วย

ภายใต้แฟ้มข้อมูลแต่ละประเภทดังกล่าว ยังประกอบไปด้วยมาตรฐานต่าง ๆ อีกเป็นจำนวนมาก โดยรูปแบบที่ใช้กับแฟ้มข้อมูล Non-Streaming Audio ที่น่าสนใจนั้น เช่น เอไอดับเบิ้ลเอฟ (Audio Interchange File Format - AIFF) ซึ่งแฟ้มข้อมูลมีนามสกุล .aif หรือ .aiff สามารถใช้ได้กับทั้งแฟ้มข้อมูลเสียงพูดและเสียงเพลง มีการทำงานเหมือนกับ ริฟเวฟ (Waveform Audio File Format - RIFF Wave) ซึ่งใช้นามสกุล .wav ทั้งนี้ เดิมที เอไอดับเบิ้ลเอฟสร้างขึ้นมาเพื่อเป็นรูปแบบเสียงมาตรฐานสำหรับระบบแมคอินทอช (Macintosh) อย่างไรก็ดี ต่อมาได้มีการสนับสนุนให้ใช้ได้กับระบบวินโดวส์ (Windows) ด้วย ในทำนองเดียวกับริฟเวฟ ที่สร้างขึ้นมาสำหรับระบบวินโดวส์ในเริ่มแรก แต่ต่อมาก็ใช้ได้กับระบบแมคอินทอชด้วย นอกจากนี้ ยังมีกฎของมิว (μ-Law) ซึ่งแฟ้มข้อมูลมีนามสกุล .au ใช้ได้กับทุกแพลตฟอร์ม แต่แฟ้มข้อมูลเสียงพูดจะมีคุณภาพสูงในขณะที่แฟ้มข้อมูลเสียงเพลงจะมีคุณภาพไม่สูงนัก รวมทั้งมิดิ (Musical Instrument Digital Interface - MIDI) ซึ่งใช้นามสกุล .mid สำหรับแฟ้มข้อมูล แฟ้มข้อมูลของมิดินั้น มีขนาดกะทัดรัดแต่ก็บรรจุได้เพียงเสียงโน้ตเท่านั้น ไม่ใช่เสียงจริง ๆ จึงมีประโยชน์สำหรับเพลงที่มีเสียงสังเคราะห์ (synthesizer-sounding music) อนึ่ง รูปแบบที่ใช้กับแฟ้มข้อมูล Non-Streaming Audio ซึ่งเป็นที่นิยมมากในปัจจุบันด้วยก็คือ เอ็มพี 3 (MPEG Audio Layer 3 - MP3) ซึ่งแฟ้มข้อมูลใช้นามสกุล .mp3 เอ็มพี3 สามารถบีบอัดได้สูงมาก และใช้ได้กับซอฟต์แวร์เครื่องเล่น (player) บนแพลตฟอร์มต่าง ๆ ส่วนรูปแบบที่ใช้กันกับแฟ้มข้อมูล Streaming Audio ซึ่งเป็นที่รู้จักนั้น เช่น QuickTime Audio โดยบริษัท Apple RealAudio โดยบริษัทRealNetworks และShockwave Audio โดยบริษัท Macromedia

สำหรับสื่อวีดิทัศน์นั้น นอกจากรูปแบบเอ็มเพ็ก (Moving Picture Experts Group - MPEG) ซึ่งมีนามสกุล .mpg หรือ .mpeg ซึ่งพบเห็นได้โดยทั่วไปแล้ว รูปแบบที่น่าสนใจสำหรับวีดิทัศน์ ที่นำมาใช้กัน คือ เอวีไอ (Audio/Video Interleaved - AVI) ซึ่งเป็นรูปแบบมาตรฐานที่คิดค้นขึ้นเพื่อใช้งานกับวีดิทัศน์สำหรับวินโดวส์ (Video for Windows - VFW) โดยแฟ้มข้อมูลมีนามสกุล .avi รวมถึงควิกไทม์มูวี่ (QuickTime Movie) ซึ่งแฟ้มข้อมูลใช้นามสกุล .mov ควิกไทม์มูวี่มีอัตราการบีบอัดที่ดี และปัจจุบันใช้กับแพลตฟอร์มต่าง ๆ ได้ ทั้งนี้ สื่อวีดิทัศน์ยังมีรูปแบบที่เรียกว่า "Streaming Video" ซึ่งเป็นไปในทำนองเดียวกับ Streaming Audio คือ แฟ้มข้อมูลเริ่มเล่นแทบจะทันทีที่มีการร้องขอโดยไม่จำเป็นต้องรอให้มีการดาวน์โหลดแฟ้มข้อมูลมาทั้งหมด และยังคงเล่นต่อไปในขณะที่มีการส่งแฟ้มข้อมูลนั้น ๆ อยู่ จึงทำให้การชมวีดิทัศน์ขนาดยาวนั้นเป็นเรื่องที่เป็นไปได้ เช่น RealVideo โดยบริษัท RealNetworks

บทสรุป

เทคโนโลยีสารนิเทศมีความสัมพันธ์กับพัฒนาการความเปลี่ยนแปลงของห้องสมุดเรื่อยมา นับตั้งแต่เริ่มมีการนำคอมพิวเตอร์เข้ามาใช้ในการดำเนินงานของห้องสมุดจวบจนถึงปัจจุบันที่มีแนวคิดเกี่ยวกับห้องสมุดดิจิทัลเกิดขึ้น โดยห้องสมุดดิจิทัลเป็นการดำเนินงานของห้องสมุดในรูปแบบใหม่และเกี่ยวข้องกับเทคโนโลยีเว็บและมาตรฐานต่าง ๆ เป็นจำนวนมาก บทความนี้ได้นำเสนอแนวคิดเรื่องห้องสมุดดิจิทัล เวิลด์ไวด์เว็บ มาตรฐานโดยรวม และมาตรฐานต่าง ๆ ที่เกี่ยวข้องสำหรับห้องสมุดดิจิทัลที่น่าสนใจ เพื่อประโยชน์ในการดำเนินงานห้องสมุดดิจิทัล

นอกจากนี้ เนื่องด้วยเทคโนโลยีสารนิเทศที่สัมพันธ์กับห้องสมุดนั้น มีการพัฒนาอยู่ตลอดเวลา บทความนี้ จึงอาจเป็นพื้นความรู้ที่ช่วยในการติดตามความเปลี่ยนแปลงต่าง ๆ เพื่อการสร้างสรรค์งานห้องสมุดดิจิทัลที่เป็นประโยชน์ และมีมาตรฐานต่อไปได้ด้วย แม้ว่าห้องสมุดดิจิทัลในอนาคตอาจเปลี่ยนแปลงไปจากที่เป็นอยู่ในปัจจุบันก็ตาม เพราะเวิลด์ไวด์เว็บและมาตรฐานที่เกี่ยวข้อง จะยังคงเป็นรากฐานที่สำคัญ ดังที่ Arms (2000: 38) กล่าวไว้ว่า ". . .Just as the crude software on early personal computers has developed into modern operating systems, the web can become the foundation for many generations of digital libraries."

รายการเว็บไซต์แนะนำที่เกี่ยวข้อง

มาตรฐาน

- American National Standards Institute (ANSI)
  http://www.ansi.org
- Encoded Archival Description (EAD) Official Web Site
  http://www.loc.gov/ead/
- Institute of Electrical and Electronics Engineers (IEEE)
  http://www.iee.org
- InterLibrary Loan Application Standards Maintenance Agency
  http://www.nlc-bnc.ca/iso/ill/
- International Organization for Standardization (ISO)
  http://www.iso.ch
- Internet Engineering Task Force (IETF)
  http://www.ietf.org

- InterNational Committee for Information Technology Standards (INCITS)
  http://www.ncits.org
- National Information Standards Organization (NISO)
  http://www.niso.org
- Text Encoding Initiative (TEI) Consortium
  http://www.tei-c.org
- Thai Industrial Standards Institute (TISI)
  http://www.tisi.go.th
- Unicode Consortium
  http://www.unicode.org
- World Wide Web Consortium (W3C)
  http://www.w3.org
- Z39.50 Maintenance Agency Page
  http://lcweb.loc.gov/Z3950/agency/

สิ่งพิมพ์

- Ariadne
  http://www.ariadne.ac.uk
- Biblio Tech Review
  http://www.biblio-tech.com
- Current Cites (Berkeley)
  http://sunsite.berkeley.edu/CurrentCites/
- D-Lib Magazine
  http://www.dlib.org
- IEEE Computer Society Digital Library Newsletter
  http://cimic.rutgers.edu/~ieeedln/
- International Journal on Digital Libraries
  http://link.springer.de/link/service/journals/00799/

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

- Journal of Digital Information
  http://jodi.ces.soton.ac.uk
- Journal of Electronic Publishing
  http://www.press.umich.edu/jep/
- The National Digital Library Periodic Reports
  http://lcweb.loc.gov/ndl/per.html
- RLG DigiNews
  http://www.rlg.org/preserv/diginews/

องค์กร

- American Society for Information Science and Technology (ASIST)
  http://www.asis.org
- Berkeley Digital Library SunSITE
  http://sunsite.berkeley.edu
- Center for the Study of Digital Libraries
  http://www.csdl.tamu.edu
- Digital Library Collaboratory Working Groups
  http://www.si.umich.edu/UMDL/EU_Grant/home.htm
- Digital Library Federation (DLF)
  http://www.clir.org/diglib/
- Digital Library Network (DLnet)
  http://www.dl.ulis.ac.jp/DLW_E/
- National Science Foundation (NSF) - Digital Libraries Initiative (DLI)
  http://www.dli2.nsf.gov
- Preserving Access to Digital Information (PADI)
  http://www.nla.gov.au/padi/
- Special Interest Group on Information Retrieval (SIGIR)
  http://www.acm.org/sigir/
- UK Office for Libraries and Networks (UKOLN)
  http://www.ukoln.ac.uk/services/elib/

สำนักวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บรรณานุกรม

ภาษาไทย

ชัยยุทธ์ ลิมลาวัลย์. Webmaster กับการบริหารเว็บไซต์. กรุงเทพฯ: ซีเอ็ดยูเคชั่น, 2544.

ต้น ตัณฑ์สุทธิวงศ์, สุพจน์ ปุณณชัยยะ, และสุวัฒน์ ปุณณชัยยะ. รอบรู้ Internet และ World Wide Web. กรุงเทพฯ: โปรวิชั่น, 2539.

ทธฤษพงศ์ เฟื่องวุฒิ. Web Design. กรุงเทพฯ: ซัม ซิสเท็ม, 2543.

ทรงลักษณ์ สกุลวิจิตร์สินธุ. "วิวัฒนาการของภาษา PostScript (ตอนที่ 2)." [ออนไลน์]. เข้าถึงได้จาก: http://www.pvinter.com/pcdirect/01/article/postscript2.htm 2544. สืบค้น 23 กรกฎาคม 2545.

ธวัชชัย ศรีสุเทพ. คัมภีร์ Web Design: คู่มือออกแบบเว็บไซต์ฉบับมืออาชีพ. กรุงเทพฯ: โปรวิชั่น, 2544.

นิรุธ อำนวยศิลป์. สร้างเว็บเพจอย่างไร้ขีดจำกัด CGI & Perl เพื่อการประยุกต์ใช้งาน: เรียนรู้ง่าย เข้าใจเร็ว พร้อมปฏิบัติได้จริง. กรุงเทพฯ: ซัคเซสมีเดีย, 2542.

พณรังสี สู่ความดี และประชา พฤกษ์ประเสริฐ. สร้างเว็บเพจอย่างไร้ขีดจำกัด ASP Active Server Pages เพื่อการประยุกต์ใช้งาน. กรุงเทพฯ: ซัคเซสมีเดีย, 2543.

พรทิพย์ โล่ห์เลขา. World Wide Web เครื่องมือใช้ Internet สำหรับทุกคน. กรุงเทพฯ: อุษาการพิมพ์, 2540.

พิธุมา พันธุ์ทวี. "ชื่อโดเมนใหม่ .aero .biz .coop .info .museum .name .pro." ไอทีปริทัศน์ 8, 11 (2543): 17-18.

สวัสดิ์ ไกรคุ้ม. การออกแบบเว็บกราฟิกด้วย HTML 4.0 ฉบับสมบูรณ์. กรุงเทพฯ: เดอะไลบรารี พับบลิชิง, 2541.

ภาษาอังกฤษ

Arms, William Y. Digital Libraries. Cambridge, MA: The MIT Press, 2000.

Baker, Thomas. "TIAC White Paper on Appropriate Technology for Digital Libraries." [Online]. Available: http://www.tiac.or.th/tiacweb/Baker/Section2_2.html 1999. Retrieved July 23, 2002.

Boss, Richard W. "Information Technology Standards." Library Technology Reports 36, 4 (2000): 1-112.

Deegan, Marilyn, and Tanner, Simon. Digital Futures: Strategies for the Information Age. London: Library Association Publishing, 2002.

Dempsey, Lorcan and others. "eLib Standards Guidelines." [Online]. Available: http://www.ukoln.ac.uk/services/elib/papers/other/standards/version2/ 1998. Retrieved June 26, 2002.

Ditto, Christopher. Webmaster: Answers! Certified Tech Support. Berkeley: Osborne/McGraw-Hill, 1998.

"EAD." [Online]. Available: http://sunsite.berkeley.edu/ead/ 2000. Retrieved June 26, 2002.

Garlock, Kristen L. and Piontek, Sherry. Building the Service-Based Library Web Site: A Step-by-Step Guide to Design and Options. Chicago: American Library Association, 1996.

Garlock, Kristen L. and Piontek, Sherry. Designing Web Interfaces: To Library Services and Resources. Chicago: American Library Association, 1999.

Haigh, Susan. "A Glossary of Digital Library Standards, Protocols, and Formats." [Online]. Available: http://www.nlc-bnc.ca/9/1/p1-253-e.html 2001. Retrieved June 26, 2002.

Interlibrary Loan Application Standards Maintenance Agency. "An Introduction to the ISO InterLibrary Loan Application Standards." [Online]. Available: http://www.nlc-bnc.ca/iso/ill/standard.htm 2002. Retrieved July 23, 2002.

International Federation of Library Associations and Institutions. "Digital Libraries: Resources and Projects." [Online]. Available: http://www.ifla.org/II/diglib.htm 2001. Retrieved June 28, 2002.

Keene, Suzanne. Digital Collections: Museums and the Information Age. Oxford: Butterworth-Heinemann, 1998.

Lynch, Patrick J., and Horton, Sarah. Web Style Guide: Basic Design Principles for Creating Web Sites. New Haven: Yale University Press, 1999.

Metz, Ray E. and Junion-Metz, Gail. Using the World Wide Web and Creating Home Pages: A How-to-Do-It Manual. New York: Neal-Schuman Publishers, 1996.

Niederst, Jennifer. Web Design in a Nutshell: A Desktop Quick Reference. Sebastopol, CA: O'Reilly, 1999.

Noerr, Peter. The Digital Library Toolkit. 2nd ed. Palo Alto, CA: Sun Microsystems, 2000.

Poulter, Alan; Tseng, Gwyneth; and Sargent, Goff. The Library Professional's Guide to the World Wide Web. London: Library Association Publishing, 1999.

"Projects Using the TEI." [Online]. Available: http://www.tei-c.org/Applications/index-az.html 2002. Retrieved July 23, 2002.

Ruth, Janice E.; Kiesling, Kris; and Spindler, Robert. "Development of the Encoded Archival Description Document Type Definition." [Online]. Available: http://lcweb.loc.gov/ead/eadback.html 1998. Retrieved July 23, 2002.

"The TEI FAQ." [Online]. Available: http://www.tei-c.org/Faq/index.html 2001. Retrieved July 23, 2002.

"The TEI Guidelines." [Online]. Available: http://www.tei-c.org/Guidelines2/index.html 2002. Retrieved July 23, 2002.

Tennant, Roy. "XML: The Digital Library Hammer." Library Journal 126, 5 (2001): 30-32.

Trin Tantsetthi. "Registration of New Charset." [Online]. Available: http://software.thai.net/alerts/tis-620/ietf-charsets.1.txt 1998. Retrieved July 26, 2002.

"The Unicode Standard: A Technical Introduction." [Online]. Available: http://www.unicode.org/unicode/standard/principles.html 2002. Retrieved July 23, 2002.

W3C Communications Team. "XML in 10 Points." [Online]. Available: http://www.w3.org/XML/1999/XML-in-10-points.html 2001. Retrieved July 23, 2002.

Waters, Donald J. "What Are Digital Libraries?" [Online]. Available: http://www.clir.org/pubs/issues/issues04.html 1998. Retrieved June 18, 2002.

"Welcome to the TEI Website." [Online]. Available: http://www.tei-c.org 2002. Retrieved June 28, 2002.

"What Is Unicode?" [Online]. Available: http://www.unicode.org/unicode/standard/WhatIsUnicode.html 2002. Retrieved July 23, 2002.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย