บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของงานวิจัย

เว็บ (Web) ถูกนำมาใช้เป็นสื่อกลางในการติดต่อสื่อสารทั้งการนำเสนอข้อมูลข่าวสารรวมถึงการทำธุรกิจต่างๆ โดยผู้ใช้บริการเว็บ (User) สามารถเข้าถึงข้อมูลที่อยู่บนเว็บโดยผ่านโปรแกรมเว็บเบราเซอร์ (Web browser) ผู้ใช้บริการเว็บนั้นยังหมายรวมไปถึงผู้มีความบกพร่องทางร่างกาย เช่น ผู้พิการทางสายตา, ผู้พิการทางการได้ยิน และอื่นๆ สำหรับผู้พิการทางสายตานั้นสามารถเข้าถึงข้อมูลของเว็บเพจโดยผ่านทางโปรแกรมอ่านหน้าจอ (Screen readers) ซึ่งโปรแกรมอ่านหน้าจอจะแปลข้อมูลในหน้าเว็บเพจให้อยู่ในรูปเสียงซึ่งแปลโดยการเรียงจากซ้ายไปขวาและจากบนลงล่างไปทีละคำ

ผู้ใช้เว็บที่มีสายตาปกติจะใช้สายตามองกวาด (Scan) ไปทั่วทั้งหน้าเว็บเพจเพื่อค้นหาข้อมูลเฉพาะที่ต้องการแทนที่จะทำการอ่านเว็บเพจอย่างละเอียด (Nielsen, 1997) สำหรับผู้พิการทางสายตาก็มีพฤติกรรมเช่นเดียวกันกับผู้มีสายตาปกติโดยใช้ความสามารถของโปรแกรมอ่านหน้าจอโดยทำการฟังอย่างรวดเร็วเพื่อค้นหาข้อมูลที่ต้องการแทนที่จะฟังข้อมูลของเว็บเพจทั้งหมด (Theofanos, 2003 และ Koyani, 2004) ตัวที่ช่วยทำให้การค้นหาเป็นไปอย่างรวดเร็วคือหัวเรื่อง (Heading) ซึ่งหัวเรื่องคือข้อมูลที่บอกให้ผู้เข้าชมเว็บเพจทราบว่ารายละเอียดของส่วนย่อย (Section) ของเว็บเพจที่จะกล่าวถึงมีเนื้อหาเกี่ยวกับอะไร และใช้ป้ายระบุหัวเรื่อง (Heading tag) เช่น <h1>, <h2>, …, <h6> ในการกำหนดรูปแบบของการแสดงผล (Mark up) การใช้หัวเรื่องไม่เพียงแต่มีประโยชน์กับผู้มีสายตาปกติแต่ยังมีประโยชน์กับผู้พิการทางตาด้วยเช่นเดียวกัน (Thatcher, 2006) ซึ่งผู้พิการทางสายตาใช้ความสามารถของโปรแกรมอ่านหน้าจอในการฟังหัวเรื่องและข้ามไปฟังอีกหัวเรื่อง หรือแสดงหัวเรื่องทั้งหมดที่มีในหน้าเว็บเพจนั้นๆ เพื่อที่ผู้พิการทางสายตาสามารถค้นหาส่วนย่อยของหน้าเว็บเพจที่มีข้อมูลที่ต้องการได้อย่างรวดเร็ว แทนที่จะต้องฟังข้อมูลทั้งหมดในหน้าเว็บเพจนั้น

ถึงแม้ว่าการใช้ป้ายระบุหัวเรื่อง (Heading tag) ในการกำหนดรูปแบบของการแสดงผลของหัวเรื่องเป็นสิ่งดี จากการศึกษาเว็บไซต์ของรัฐบาลไทยจำนวน 200 เว็บไซต์กลับพบว่ามีเพียง 11 เปอร์เซ็นต์เท่านั้นที่มีการใช้ป้ายระบุหัวเรื่องในเว็บไซต์ และ (Watanabe, 2007) ได้ศึกษาว่าผู้พิการทางสายตาได้รับผลกระทบมากกว่าผู้มีสายตาปกติในการใช้งานเว็บเพจที่หัวเรื่องไม่ได้ใช้

ป้ายระบุหัวเรื่องในการกำหนดรูปแบบการแสดงผล เนื่องจากผู้มีสายตาปกติสามารถเข้าใจและระบุหัวเรื่องจากเว็บเพจได้จากรูปแบบการแสดงผลแม้ว่าหัวเรื่องนั้นจะไม่ได้ใช้ป้ายระบุหัวเรื่องในการกำหนดรูปแบบการแสดงผลก็ตาม เช่น หัวเรื่องจะมีขนาดตัวอักษรที่ใหญ่กว่าตัวอักษรอื่นๆในส่วนย่อยของเว็บเพจนั้น (Watanabe, 2007) แต่กับผู้พิการทางสายตาไม่สามารถทำเช่นเดียวกันกับผู้มีสายตาปกติได้เนื่องจากโปรแกรมอ่านหน้าจอจะสามารถระบุว่าส่วนใดเป็นหัวเรื่องได้ก็ต่อเมื่อหัวเรื่องนั้นต้องถูกกำหนดรูปแบบการแสดงผลด้วยป้ายระบุหัวเรื่องเท่านั้น

จากปัญหาที่กล่าวมาข้างต้น วิทยานิพนธ์นี้จะนำเสนอการใช้เทคนิคของการแบ่งเว็บเพจเป็นส่วนย่อยด้วยวีไอพีเอสอัลกอริทึมควบคู่กับการใช้ซัพพอร์ตเวกเตอร์แมชชีนซึ่งเป็นวิธีการทางด้านการเรียนรู้ด้วยเครื่องจักร(Machine learning) มาทำการจำแนก (Classification) โหนดข้อความซึ่งได้จากดอม (DOM: Document Object Model) ว่าเป็นหัวเรื่องหรือไม่จากรูปแบบการแสดงผล โดยจะเปรียบเทียบรูปแบบการแสดงผลของโหนดข้อความนั้นเปรียบเทียบกับโหนดข้อความอื่นๆภายในส่วนย่อยของเว็บเพจนั้น ได้แก่ ขนาดตัวอักษรของข้อความ, สีตัวอักษรของข้อความ, สีพื้นหลังอักษรของข้อความ, ไอคอน (Icon) หน้าข้อความ, การย่อหน้าของข้อความ และตำแหน่งของเอชทีเอมแอลของข้อความภายในส่วนย่อยนั้น

ผลการศึกษาที่ได้จากงานวิจัยนี้สามารถนำไปประยุกต์ใช้ในเพิ่มความสามารถให้กับโปรแกรมอ่านหน้าจอ ในส่วนของการตรวจจับหัวเรื่องที่ไม่ได้ใช้ป้ายระบุหัวเรื่องในการกำหนดรูปแบบการแสดงผล ซึ่งเป็นการเอื้อประโยชน์ต่อผู้พิการทางสายตาให้สามารถฟังหัวเรื่องและข้ามไปฟังอีกหัวเรื่องเพื่อค้นหาข้อมูลที่ต้องการได้แม้ว่าหัวเรื่องนั้นจะไม่ได้ถูกสร้างขึ้นด้วยป้ายระบุหัวเรื่องก็ตาม นอกจากนั้นยังสามารถนำไปประยุกต์ใช้กับโปรแกรมตรวจสอบความถูกต้องของการสร้างเว็บไซต์ตามมาตราฐานการออกแบบเว็บไซต์เพื่อให้ทุกคนเข้าถึงได้ (Wave, Site Valet, HiSoftware) ว่ามีการใช้หัวเรื่องแต่ไม่ได้ใช้ป้ายระบุหัวเรื่องในการกำหนดรูปแบบการแสดงผลเพื่อที่ผู้ออกแบบเว็บเพจจะทำการแก้ไขเพื่อใช้ป้ายระบุในการกำหนดรูปแบบการแสดงผลที่ถูกต้องตามวัตถุประสงค์

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อศึกษาแนวทางวิจัย และกระบวนการในการพัฒนาวิธีการจำแนกหัวเรื่องจากเว็บเพจจากรูปแบบการแสดงผล

2. เพื่อเป็นแนวทางในการประยุกต์ใช้เพื่อเพิ่มความสามารถให้กับโปรแกรมอ่านหน้าจอ โปรแกรมแปลงเว็บเพจเพื่ออุปกรณ์ที่มีหน้าจอขนาดเล็ก และอื่นๆ ให้มีความสามารถในการจำแนกหัวเรื่องที่ไม่มีป้ายระบุหัวเรื่องในการกำหนดรูปแบบการแสดงผล
3. เพื่อเสนอแนวทางในการศึกษาและพัฒนางานวิจัยต่อเนื่องและข้อเสนอแนะในการพัฒนาวิธีการจำแนกและตรวจสอบการใช้ป้ายระบุเอชทีเอมแอลให้ถูกต้องตามความหน้าที่ที่ป้ายระบุนั้นใช้กำหนดรูปแบบการแสดงผล

1.3 ผลที่ได้รับ

1. วิธีการจำแนกหัวเรื่องที่ไม่ได้ใช้ป้ายระบุหัวเรื่องในการกำหนดรูปแบบการแสดงผล เพื่อที่ผู้พิการทางสายตาสามารถท่องเว็บไซต์ได้สะดวกขึ้นจากความสามารถในการจำแนกหัวเรื่องที่ไม่ได้ใช้ป้ายระบุหัวเรื่องได้
2. แนวทางการจำแนกหัวเรื่องที่ไม่ได้ใช้ป้ายระบุหัวเรื่อง (Heading tag) เพื่อนำไปพัฒนาความสามารถโปรแกรมอ่านหน้าจอ, โปรแกรมแปลงเว็บเพจเพื่ออุปกรณ์ที่มีหน้าจอขนาดเล็ก
3. แนวทางในการศึกษาและนำไปประยุกต์ใช้กับการจำแนกและตรวจสอบการใช้ป้ายระบุเอชทีเอมแอล ประเภทอื่นๆ ให้ถูกต้องตามรูปแบบการแสดงผลของป้ายระบุเอชทีเอมแอล

1.4 รายละเอียดของวิทยานิพนธ์

รายละเอียดของวิทยานิพนธ์ในส่วนที่เหลือสามารถแบ่งออกเป็นส่วนต่าง ๆ ซึ่งสามารถแยกอธิบายได้ดังหัวข้อต่อไปนี้

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ประกอบด้วยทฤษฎีต่างๆ ที่เกี่ยวข้องกับงานวิจัยนี้ งานวิจัยที่ประยุกต์ใช้เทคนิคต่างๆ กับการสรุปความจากเอกสาร และงานวิจัยที่ใช้เทคนิคการค้นหาเส้นทางที่สั้นที่สุดประยุกต์ใช้กับการสรุปความจากเอกสาร

บทที่ 3 วิธีการดำเนินงานวิจัย ประกอบด้วยรายละเอียดของข้อมูลที่ใช้ในการวิจัย เครื่องมือที่ใช้ ขั้นตอนของการวิจัย และวิธีวัดผลการวิจัย

บทที่ 4 ผลการทดลอง แสดงผลการทดลองที่ได้แต่ละกระบวนการ

บทที่ 5 สรุปผลการศึกษาและข้อเสนอแนะ