

การพัฒนาระบบบริการตรวจสอบและประมวลผลสถิติการเยี่ยมชมเว็บไซต์ในประเทศไทย (<http://truehits.net>)

จารุวรรณ ละอองมัลย์ สิทธิชัย ลีลานั้นท์นุกูล ปิยะ ดัฒทวีเชียร

สำนักบริการเทคโนโลยีสารสนเทศภาครัฐ (สทสร.)

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

ชั้น 17 อาคารบางกอกไทยทาวเวอร์ ถนนรางน้ำ เขตราชเทวี กรุงเทพฯ 10400

webmaster@truehits.net

ABSTRACT - This paper presents a promising web visitor monitoring system named "TrueHits - <http://truehits.net>" that gathers information from web sites every time visitors visit them. Such information shows where, when and how visitors access a particular web site. This information is then processed and analyzed for giving us the useful statistics. Those statistics are very useful not only to the web site owner but also to the public. For the web site owner, it is useful for the development of web content to make it more suitable to their target visitors based on their behavior. For the country's big picture, the statistics show Internet usage behavior of people in Thailand. The TrueHits statistics shown in this paper is obtained by processing and analyzing the gather information totally 2,354,012,488 data records from January 1, 2002 to December 31, 2002.

บทคัดย่อ - บทความฉบับนี้นำเสนอการวิจัยพัฒนาระบบให้บริการตรวจสอบและประมวลผลสถิติการเยี่ยมชมเว็บไซต์ในประเทศไทย หรือที่เรียกว่า ทรูฮิต (TrueHits - <http://truehits.net>) ซึ่งมีประโยชน์ต่อผู้พัฒนาเว็บไซต์สามารถนำไปใช้เป็นเครื่องมือวัดในการปรับปรุงเนื้อหาเว็บไซต์ให้ดียิ่งขึ้นและให้เหมาะกับกลุ่มเป้าหมายที่เข้ามาใช้งานด้วย และมากกว่านั้น ได้รวบรวมสถิติการเยี่ยมชมเว็บไซต์เมืองไทยที่น่าสนใจ ซึ่งสะท้อนให้เห็นพฤติกรรมภาพรวมของประเทศไทยในการใช้อินเทอร์เน็ตได้เป็นอย่างดี สถิติที่ได้เกิดจากการประมวลผลและวิเคราะห์ข้อมูลของผู้เข้าเยี่ยมชมเว็บไซต์โดยตรงจากการใช้งานจริงของเว็บไซต์ภายในประเทศที่รวบรวมได้จากจำนวนเว็บไซต์สมาชิก ทรูฮิต ในระหว่างวันที่ 1 มกราคม พ.ศ. 2545 ถึง 31 ธันวาคม พ.ศ. 2545 โดยมี จำนวนข้อมูลดิบทั้งหมดที่นำมาวิเคราะห์ถึง 2,354,012,488 ชุดข้อมูล

1. บทนำ

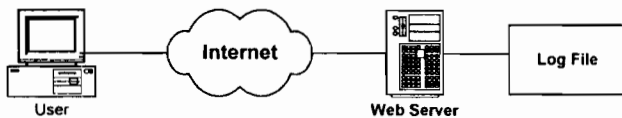
ปัจจุบันประเทศไทยมีการใช้อินเทอร์เน็ต (Internet) กันอย่างแพร่หลาย และมีแนวโน้มการเติบโตของจำนวนประชากรอินเทอร์เน็ตในประเทศไทยเพิ่มขึ้นอย่างรวดเร็ว กิจกรรมหลักอย่างหนึ่งที่ได้รับความนิยมมากในการใช้งานผ่านอินเทอร์เน็ตก็คือการให้บริการ World Wide Web (WWW) เพื่อเข้าถึงข้อมูลต่างๆ ที่อยู่บนเว็บไซต์ โดยอาศัยหลักการทำงานของโปรโตคอล Hypertext Transfer Protocol (HTTP) ส่งผลให้มีเว็บไซต์เกิดขึ้นมากมาย เว็บไซต์บางเว็บไซต์ได้รับความนิยมจากผู้ใช้อินเทอร์เน็ตเป็นจำนวนมาก โดยมีกรกล่าวกันว่าจำนวนครั้งที่ผู้เข้า

เยี่ยมชมเข้ามาที่เว็บไซต์เฉลี่ยแล้ววันหนึ่งๆ เกินหนึ่งล้านครั้ง ในขณะที่เว็บไซต์อีกจำนวนไม่น้อยไม่ได้รับความนิยมเท่าที่ควร วิธีการวัดความนิยมของเว็บไซต์นั้นโดยทั่วไปจะหมายถึงการวัดสถิติการเยี่ยมชมเว็บไซต์ซึ่งมีวิธีตรวจสอบได้หลายวิธี ดังต่อไปนี้

1. ติดตั้งโปรแกรมเคาน์เตอร์ [1] (Web Counter) เพื่อบันทึกจำนวนครั้ง (click) ของการที่มีผู้เข้ามาที่หน้าเว็บไซต์ คือจะมีตัวเลขที่แสดงจำนวนครั้งที่มีการเยี่ยมชม วิธีนี้มีข้อเสียคือค่าเคาน์เตอร์จะเป็นค่าที่นับสะสมไปเรื่อยๆ ไม่สามารถแสดงสถิติการเยี่ยมชมเทียบกับช่วงเวลาชั่วโมงวัน เดือนปีได้ กล่าวคือเป็นแค่การนับปริมาณ click ที่มีการเก็บเข้ามา ยก

ตัวอย่างเช่น มีคนคนหนึ่ง click หน้าเว็บจำนวน 100 ครั้ง ตัวเลขเคาน์เตอร์ก็จะนับเพิ่มเป็น 100 ใดๆ ที่เป็นคนเดียวที่เข้ามาเยี่ยมชม

2. Web Log File Analysis [2] หรือ Local Monitoring วิธีนี้ผู้ดูแลเว็บไซต์ (Web Admin) ทำการตรวจสอบข้อมูลสถิติของเว็บไซต์ของตนเองได้จาก Log file ของ Web Server โดยมีหลักการทำงานอธิบายได้ดังแสดงในรูปที่ 1



รูปที่ 1. แสดงรูปแบบของการตรวจสอบสถิติแบบ Web Log File Analysis

เมื่อมี User ทำการ request เว็บเพจมายังเว็บไซต์ เว็บเซิร์ฟเวอร์จะเก็บข้อมูลลงใน Log File เรียงตามเวลาคล้ายตัวอย่างเช่น

```
hostname - - [01/Feb/1998:10:10:00 +0100] "GET /index.html
HTTP/1.0" 200 4839
```

ผู้ดูแลเว็บไซต์จะเขียนโปรแกรม (หรือ script) หรือติดตั้งโปรแกรมสำเร็จรูปเพื่อทำการวิเคราะห์ข้อมูลจาก Log File ดังกล่าวให้เป็นข้อมูลในรูปแบบรายงานที่อ่านเข้าใจง่ายและสามารถดูผ่าน WWW ได้ ซึ่งข้อเสียวิธีนี้คือจำเป็นต้องติดตั้งดูแลโปรแกรมวิเคราะห์ซึ่งต้องอาศัยความเชี่ยวชาญรวมทั้งอาจเกิดค่าใช้จ่ายในการจัดซื้อโปรแกรมวิเคราะห์ถ้าพวก Commercial Software และเนื่องจากการวัดสถิติโดยเจ้าของเว็บไซต์เองดังนั้นจึงมีปัญหาเรื่องความน่าเชื่อถือของสถิติที่วัดได้ รวมทั้งสถิติดังกล่าวไม่ได้เป็นมาตรฐานเดียวกันทั้งหมด จึงไม่สามารถนำมาเปรียบเทียบสถิติแต่ละเว็บไซต์ได้ จากข้อจำกัดของทั้งวิธีการวัดสถิติโดยใช้เคาน์เตอร์และ Web Log Analysis จึงได้มีการวัดสถิติการเยี่ยมชมเว็บไซต์ ที่ให้บริการแบบ Remote Monitoring

3. Remote Monitoring [3] คือผู้ดูแลเว็บไซต์ไม่จำเป็นต้องเก็บ Log File การใช้งานของ Web Server เอง แต่ใช้บริการวัดสถิติจากผู้ให้บริการที่ได้รับการยอมรับจัดการเก็บข้อมูล ประมวลผล วิเคราะห์การใช้งานเว็บไซต์แทนและแสดงผลในรูปแบบของรายงานให้ด้วย ในปัจจุบันมีผู้ให้บริการในต่างประเทศหลายแห่ง ได้แก่ hitbox [4], nedstat [5], extreme tracking [6], stat4all [7] เป็นต้น ซึ่งได้รับความนิยมมากจาก

เว็บไซต์ต่างๆ รวมทั้งจากเว็บไซต์ในประเทศไทยเนื่องด้วยความสะดวกและง่ายในการใช้บริการ อีกทั้งการแสดงสถิติการเยี่ยมชมสามารถเปรียบเทียบตามช่วงเวลาได้

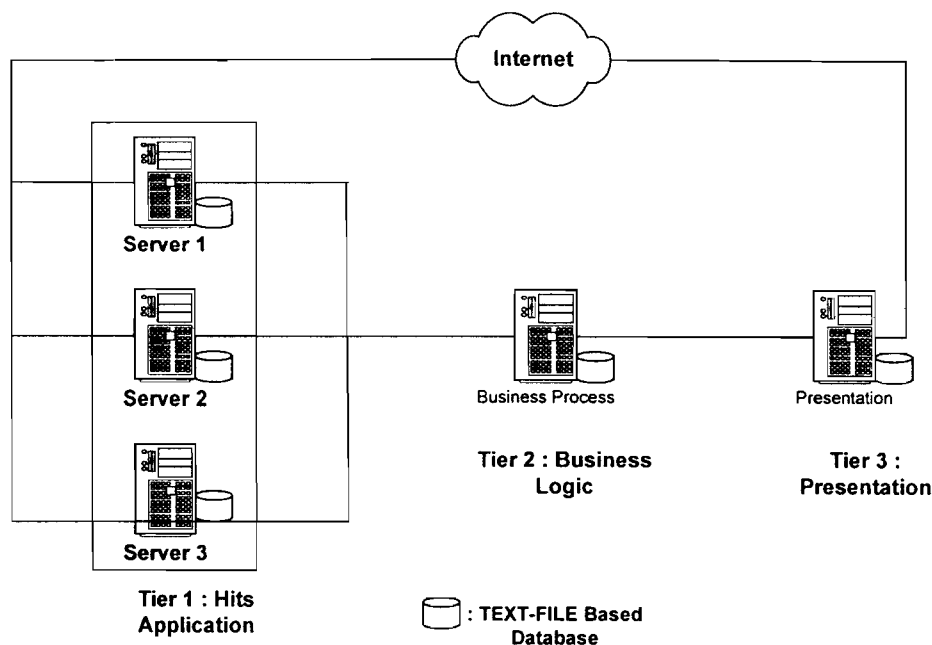
อย่างไรก็ตามระบบให้บริการตรวจสอบสถิติการเยี่ยมชมเว็บไซต์แบบ Remote Monitoring ที่สมบูรณ์และมีคุณภาพนั้น จะเป็นผู้ให้บริการของต่างประเทศเป็นส่วนใหญ่ เว็บไซต์ของหน่วยงานต่างๆ ในประเทศไทยจำเป็นต้องพึ่งพาระบบดังกล่าวของต่างประเทศเป็นจำนวนมากส่งผลให้ข้อมูลการเยี่ยมชมเว็บไซต์ไทยอยู่ในมือของผู้ให้บริการในต่างประเทศ ดังนั้นนักวิจัยไทย จึงไม่สามารถนำข้อมูลเหล่านั้น มาประมวลผลเป็นสถิติที่มีประโยชน์ได้ ยิ่งไปกว่านั้น นอกจากข้อมูลสถิติที่ได้รวบรวมขึ้นจะเป็นประโยชน์ในการวิเคราะห์ความนิยมเว็บไซต์แล้ว ยังจะกลายเป็นเครื่องมือหนึ่งที่ช่วยส่งเสริมให้เกิดการพัฒนาและปรับปรุงเนื้อหาในเว็บไซต์ให้ดียิ่งขึ้น และเป็นเครื่องมือวัดที่มีมาตรฐานเดียวกันด้วย

ในปัจจุบันนอกจากมีการวัดสถิติการเยี่ยมชมเว็บไซต์ใน 3 แบบดังกล่าวข้างต้นแล้ว ยังมีวิธีการทำแบบสอบถาม Online ซึ่งในประเทศไทยสำนักงานเลขาธิการคณะกรรมการเทคโนโลยีสารสนเทศแห่งชาติ NITC ทำวิจัยเพื่อสำรวจกลุ่มผู้ใช้อินเทอร์เน็ตโดยใช้แบบสอบถาม [8] หนึ่งในข้อจำกัดของวิธีใช้แบบสอบถามคือ ไม่สามารถดูการใช้งานได้ตลอดทุกช่วงเวลา ดังนั้น ระบบทฤษฎีจึงถูกพัฒนาขึ้นเพื่อใช้เป็นเครื่องมือที่ใช้จัดเก็บข้อมูลการใช้งานตั้งแต่ปี พ.ศ. 2545 จนถึงปัจจุบัน และวิเคราะห์ถึงพฤติกรรมการใช้อินเทอร์เน็ตภาพรวมของเมืองไทยได้ตลอดเวลา

2. แนวคิดการออกแบบและหลักการทำงานของระบบทฤษฎี

2.1 แนวคิดการออกแบบระบบ

แนวคิดหลักในการออกแบบระบบทฤษฎีเพื่อใช้เป็นเครื่องมือในการให้บริการตรวจสอบสถิติการเยี่ยมชมเว็บไซต์ของประเทศไทย ภายใต้เงื่อนไขที่ต้องสามารถรองรับปริมาณทราฟฟิกได้จำนวนสูงมาก ไม่น้อยกว่า 15 ล้าน transactions/วัน และ ประมวลผลเพื่อสรุปสถิติการเยี่ยมชมต่อเว็บสมาชิกทฤษฎีทั้งหมดทุกๆ ชั่วโมง โดยต้องออกแบบโครงสร้างของระบบเพื่อรองรับการเจริญเติบโตของปริมาณทราฟฟิก, ปริมาณข้อมูล (พื้นที่ Storage ที่ใช้เก็บข้อมูล) และ จำนวนสมาชิกที่เพิ่มขึ้นอย่างต่อเนื่อง และสามารถให้บริการได้ตลอดเวลา (Non Stop Service)



รูปที่ 2. แสดงโครงสร้างการออกแบบระบบทูลิต

จากรูปที่ 2 เห็นว่าโครงสร้างดังกล่าวออกแบบไว้เป็น Multi Tier Architecture [9] คือ

1. Tier 1: Hits Application คือชั้นของเครื่องคอมพิวเตอร์แม่ข่ายที่ทำหน้าที่รับข้อมูลสถิติ จากเว็บไซต์ของสมาชิกทั้งหมด ซึ่งเครื่องจะมี Load การทำงานที่สูงเพราะต้องรองรับ Concurrent Transaction พร้อม ๆ กัน ได้โดยปัจจุบันทำการ Setting ไว้ที่ 1024 Concurrent Transaction เนื่องจากระบบออกแบบเพื่อประโยชน์ในด้าน Cost-Effective ดังนั้นจึงใช้เทคนิครับข้อมูลสถิติแบบกระจายเครื่องโดยดูว่าเว็บไซต์ที่มีข้อมูลสถิติมาก จะใช้เครื่องที่รับสถิติที่มีประสิทธิภาพสูงร่วมกัน และเว็บไซต์ที่มีข้อมูลสถิติไม่มาก จะใช้เครื่องรับสถิติที่มีประสิทธิภาพไม่สูงนักร่วมกัน และฐานข้อมูลที่เก็บข้อมูลรับสถิติ เราออกแบบให้เป็น Text-Based ซึ่งมีข้อดีกว่าการใช้ Database คือ สามารถทำงานได้เร็วกว่า และพื้นที่ที่จัดเก็บใน File น้อยกว่าเก็บใน Database เนื่องจาก Database ต้องเปลืองพื้นที่ในส่วน of Overhead ค่อนข้างมาก

2. Tier 2: Business Login Tier คือชั้นของเครื่องคอมพิวเตอร์แม่ข่ายที่ทำหน้าที่ประมวลผล เนื่องจากต้องการความเร็วในการประมวลผลมาก ดังนั้นทูลิตจึงพัฒนาโปรแกรมที่ใช้คำนวณให้ทำงานได้เร็ว โดยพัฒนาด้วยภาษา C เพื่อคำนวณข้อมูลประมาณ 15 ล้าน transactions/วัน ได้รวมทั้งต้องประมวลผลข้อมูลทั้งเดือนได้ คือประมาณ 15 ล้าน transactions*30 วัน ต่อเดือน คือประมาณ 450 ล้าน transactions ใน 1 เดือนได้เพื่อเก็บเป็นสถิติภาพรวมรายเดือน

เนื่องจากจำนวนสมาชิกในช่วงปี 2545 มีประมาณ 7000 เว็บไซต์ ทูลิตได้ทำการประมวลผลแบบ Parallel Processing คือทำงานขนานกันไป จะทำได้ให้ข้อมูลสถิติโดยใช้เวลาการประมวลผลที่เร็วขึ้น

3. Tier 3: Web Presentation Tier คือชั้นของเครื่องคอมพิวเตอร์ที่ทำหน้าที่แสดงผลข้อมูลสถิติ ซึ่งต้องการเนื้อที่ในการจัดเก็บข้อมูลอย่างมาก เพราะระบบทูลิตเก็บข้อมูลสถิติย้อนหลังตลอดช่วงเวลาการให้บริการของทุกเว็บไซต์โดยที่ระบบ จะ Generate ผลสรุปสถิติให้อยู่ในรูปแบบ HTML File เพื่อความรวดเร็วในการเรียกดูข้อมูลสถิติ ดังนั้นนอกจากการใช้วิธีการเพิ่ม Disk Storage แล้วทูลิตยังทำด้าน โปรแกรมโดยทำการบีบอัดข้อมูลโดยใช้ bzip2 [10] ทำให้ข้อมูลที่เก็บมีขนาดลดลงประมาณ 80 เปอร์เซ็นต์เนื่องจากข้อมูลเป็นข้อมูลประเภท Text File จึงทำให้บีบอัดข้อมูลได้มากและในขณะที่เรียกดูข้อมูลนั้นจึงทำการ Unzip ข้อมูล ซึ่งช่วยประหยัดพื้นที่ในการจัดเก็บข้อมูลได้มาก

จากการออกแบบข้างต้น ข้อดีที่เห็นได้ชัดเจน คือ

1. Cost Effective คือใช้เครื่องคอมพิวเตอร์แม่ข่ายที่ใช้ CPU สนับสนุนสถาปัตยกรรม X86 และใช้ระบบปฏิบัติการลินุกซ์ทั้งหมดในการให้บริการได้ตลอดเวลา
2. Scalability สามารถทำการขยายเครื่องรับข้อมูลสถิติ เครื่องแสดงผล และเครื่องประมวลผลได้ตามที่ต้องการ โดยไม่กระทบกับระบบเดิมที่ให้บริการอยู่
3. Simplicity ออกแบบระบบง่าย และใช้งานได้จริงในทางปฏิบัติ (Practical Usage)

2.2 หลักการทำงานของระบบทรูฮิต

หลักการทำงานของทรูฮิตนั้นอาศัยการทำงานของโปรโตคอล HTTP (Hypertext Transfer Protocol) [11] ซึ่งมีขั้นตอนการทำงานโดยสังเขปดังนี้ เริ่มจาก client ซึ่งใช้โปรแกรม web browser ร้องขอบริการ HTTP ผ่าน TCP Connection ไปยัง Web Server ซึ่งภายหลังการเชื่อมต่อระหว่าง Web Server และ client สำเร็จแล้ว เครื่อง client จะส่ง Request ไปยัง Web Server เมื่อ Web Server ได้รับ request แล้วจะส่ง Response กลับไปให้ Client จากนั้น หลังจากการส่งข้อมูลเสร็จสิ้น Web Server จะทำการ Close Connection ทั้งนี้ ในส่วนของการส่ง request ของ client และการส่ง response กลับของ server สามารถแสดงรายละเอียดดังต่อไปนี้

1. The HTTP Request

วิธีที่ client ใช้ในการ Request นั้นมีหลายวิธีด้วยกัน วิธีที่มักใช้งานโดยทั่วไป คือ GET, POST และ PUT ขอยกตัวอย่างในกรณีของการใช้ GET method ที่อยู่ในรูปแบบของ HTTP Request Format

```
GET http://truehits.net/ HTTP/1.0
```

2. HTTP Response Format

หลังจากที่ Server ได้รับ Request มาจาก Client แล้วนั้น จะส่ง Response อยู่ใน Format ดังต่อไปนี้

```
HTTP/1.1 200 OK
```

```
Date: Thu, 27 Feb 2003 09:15:21 GMT
```

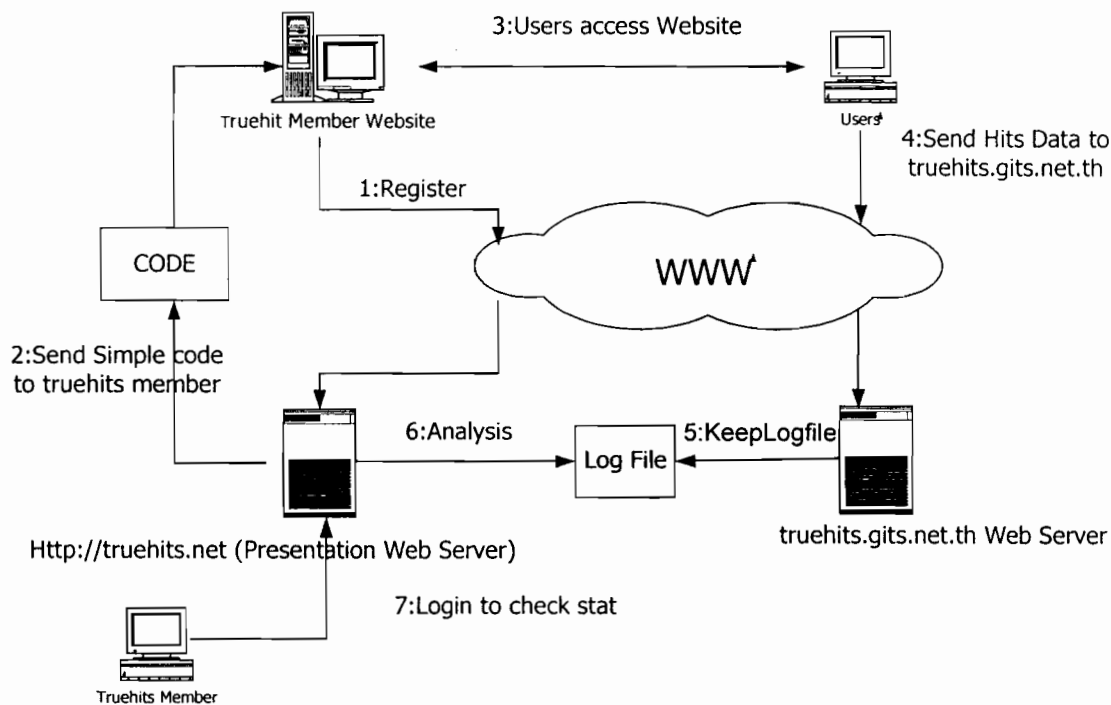
```
Server: Apache
```

```
Connection: close
```

```
Content-Type: text/html
```

```
..... <ข้อมูล html> .....
```

ซึ่งระบบการทำงานของทรูฮิตอาศัย HTTP Protocol ดังข้างบนและมีหลักการทำงานดังแสดงในรูปที่ 3



รูปที่ 3. แสดงรูปแบบของการตรวจสอบสถิติแบบ Remote Monitoring ของระบบทรูฮิต

จากรูปที่ 3 สามารถอธิบายขั้นตอนการทำงาน ได้ดังนี้

1. สมาชิกทรูฮิตทำการสมัครใช้งาน ดังหมายเลข 1
2. หลังจากสมัครแล้ว ระบบทรูฮิตส่งโค้ดไปให้สมาชิก ดังหมายเลข 2 โดยโค้ดที่ส่งไปมีตัวอย่าง คือ

```
<SCRIPT
LANGUAGE="javascript1.1"src="http://truehits.gits.net.th/data/a0000
007.js">
</SCRIPT>
```

ซึ่งตัวอย่างโค้ด JavaScript ของสมาชิก 1 เว็บไซต์แสดงได้ดังต่อไปนี้

```
var truehitsurl=document.URL;
var page;
var truehitspage=page;
var turlnameindex="tmtelephone.com";
var truehitsflag;
if ( (!truehitsflag) && ( truehitsurl.indexOf(turlnameindex) > 0) )
{
rf=escape(document.referrer);
if((rf=="undefined")||(rf=="")){rf="bookmark";};
bn=navigator.appName;if(bn.substring(0,9)=="Microsoft"){bn="MSIE";};
bv=Math.round(parseFloat(navigator.appVersion)*100);
sv=1.1;ja=(navigator.javaEnabled()===true)?"y":"n";
ck="";var cookiecx=(new Date()).toGMTString();
document.cookie="verify=test; expire="+cookiecx;
if(document.cookie.length>0){ck="y";} else {ck="n";}
document.write("<script language=javascript1.2>");
document.write("sv=1.2;ss=screen.width+'*'+screen.height;sc=(bn=='MSIE')?screen.colorDepth:screen.pixelDepth;if(sc=='undefined'){sc='na';}");
document.write("</"+>script");
document.write("<script language=javascript1.3>");
document.write("sv=1.3;");
document.write("</"+>script");
arg+"&bn="+bn+"&bv="+bv+"&ss="+ss+"&sc="+sc+"&sv="+sv+"&ck="+ck;
arg+="&ja="+ja+"&rf="+rf;
x1="<a href='http://truehits.net/stat.php?login=tmtelephone' target='_blank'>";
x2="<img src='http://truehits.gits.net.th/gengif.php3?hc=a0000007';";
x3=" width=16 height=17 border=0";
x4=" alt='truehits-logo'></a>";
```

```
document.write(x1+x2+arg+"&truehitspage="+truehitspage+"&truehitsurl="+truehitsurl+x3+x4);
truehitsflag=1;
}
```

- หลังจากที่สมาชิกนำโค้ด JavaScript ไปใส่ไว้ที่เว็บไซต์ของตนเองแล้วจะเห็นรูปภาพขึ้นที่หน้านั้น เมื่อมีการเรียกหน้าเพจ ดังหมายเลข 3 code ทฤษฎีจะส่งข้อมูลของ user ไปยัง truehits.gits.net.th Web Server ดังหมายเลข 4 ซึ่งระบบจะแยกข้อมูลของแต่ละเว็บไซต์เป็นคนละ log file เพื่อง่ายต่อการวิเคราะห์ข้อมูลสถิติต่อไป ดังหมายเลข 5
- ระบบทฤษฎีทำการ Analyze ข้อมูล ดังหมายเลข 6 และสมาชิกสามารถ login เข้ามาดูข้อมูลสถิติของตัวเองได้ที่ <http://truehits.net> โดยมีรูปแบบ report ที่อ่านเข้าใจง่ายผ่านทาง WWW ดังหมายเลข 7

ระบบ TrueHits Web Server จะเก็บข้อมูลต่างๆ ของ User ที่เข้าชมเว็บไซต์ลงในข้อมูล Text File ซึ่งมีตัวอย่างของ Format ของ 1 ชุดข้อมูล (1 record) ดังนี้ คือ

```
MSIE 2003/02/25 20:02:51 Mozilla/4.0 (compatible; MSIE 5.5;
Windows 98; Win 9x 4.90) 202.28.169.165 800*600 16 1.3 y
http://www.google.co.th/search?q=%E0%B8%99%E0%B8%B4%E0%B9%80%E0%B8%A7%E0%B8%A8%E0%B8%A7%E0%B8%B4E0%B8%97%E0%B8%A2%E0%B8%B2%E0%B8%97%E0%B8%B2%E0%B8%87%E0%B8%97%E0%B8%B0%E0%B9%80%E0%B8%A5&ie=UTF-8&oe=UTF-8&hl=th th
http://webindex.sanook.com/science_and_engineering/chemistry_and_biology/ y Sticker Page 1.0 proxy.mahidol.ac.th:8080 (Squid/2.2.STABLE5) 191.1.3.42 fe80859ba60b2de49f70fac854c4234a
```

ซึ่งรายละเอียดของ field ต่างๆ ที่ถูกจัดเก็บใน Text File สามารถแยกตามข้อตัวแปรได้ดังตารางที่ 1

ตารางที่ 1. แสดงความสัมพันธ์ระหว่างค่าที่ได้และชื่อตัวแปรจากข้อมูลสถิติ 1 ชุดข้อมูล

ลำดับ	ค่าที่ได้	ชื่อตัวแปร
1	MSIE	Browser Name
2	2003/02/25 20:02:51	Date time
3	Mozilla/4.0 (compatible; MSIE 5.5; Windows 98; Win 9x 4.90)	\$HTTP_USER_AGENT
4	202.28.169.165	\$REMOTE_ADDR
5	800*600	Resolution
6	16	Color Depth
7	1.3	Java Script Version
8	y	Java Enable
9	http://www.google.co.th/search?q=%E0%B8%99%E0%B8%B4%E0%B9%80%E0%B8%A7%E0%B8%A8%E0%B8%A7%E0%B8%B4%E0%B8%97%E0%B8%A2%E0%B8%B2%E0%B8%97%E0%B8%B2%E0%B8%87%E0%B8%97%E0%B8%B0%E0%B9%80%E0%B8%A5&ie=UTF-8&oc=UTF-8&hl=th%u0E0B%u0E32%u0E27%u0E14%u0E4C	Referrer URL
10	Th	\$HTTP_ACCEPT_LANGUAGE
11	http://webindex.sanook.com/science_and_engineering/chemistry_and_biology	document.URL
12	Y	Cookie Enable
13	Sticker Page	Page Name
14	1.0 proxy.mahidol.ac.th:8080 (Squid/2.2.STABLE5)	\$HTTP_VIA
15	191.1.3.42	\$HTTP_X_FORWARDED_FOR
16	fe80859ba60b2de49f70fae854c4234a	Session ID

ตัวแปรแต่ละตัวแปรมีความหมายดังต่อไปนี้

- 1) Browser Name คือ ชื่อของเว็บ Browser
- 2) Date time เป็นวันและเวลาในประเทศไทย
- 3) \$HTTP_USER_AGENT เป็น CGI Environment variable [12] ที่บอกชนิดของ Browser ที่ใช้ในการเข้าถึง CGI ทางฝั่ง Web Server
- 4) \$REMOTE_ADDR เป็น CGI Environment Variable ที่ระบุ IP ของ Client
- 5) Resolution คือ ความละเอียดของหน้าจอ
- 6) Color Depth คือจำนวนบิตสีที่เซดที่หน้าจอ Web Browser
- 7) JavaScript Version คือ Version ของ JavaScript มีตั้งแต่ 1.0, 1.1, 1.2 และ 1.3
- 8) Java Enable ตรวจสอบว่าที่ Web Browser Enable Java หรือไม่

- 9) Referrer URL คือ URL ก่อนหน้านั้น ที่จะเข้ามาเรียกที่เว็บไซต์หน้าปัจจุบัน ยกตัวอย่าง เช่น เว็บ A มี link เว็บ เป็น `<a href="http://webb.com">webb</a>` ถ้ากดที่ link นั้นเข้ามายังเว็บ webb.com จะบอกว่ามี Referrer มาจากเว็บ A เป็นต้น
- 10) \$HTTP_ACCEPT_LANGUAGE เป็น CGI Environment Variable ที่บ่งบอกภาษาที่กำหนดไว้ที่ Web Browser ซึ่งสามารถดูชื่อเต็มภาษาของตัวย่อได้จาก `http://www.nectec.or.th/users/htk/public/iso-country-code`
- 11) document.URL คือ หน้า URL ปัจจุบัน เช่น URL คือ sanook.com เมื่อเข้าที่เว็บ sanook.com
- 12) Cookie Enable ตรวจสอบว่า Web Browser enable cookie หรือไม่
- 13) Page Name คือการที่เจ้าของเว็บไซต์สามารถที่จะกำหนดชื่อหน้าได้เอง เพื่อใช้ตรวจสอบว่าหน้าไหนมีปริมาณคนเข้าเยี่ยมชมมากน้อยเพียงไร

- 14) SHTTP_VIA เป็น CGI Environment Variable ที่บอกชื่อของ Proxy Server
- 15) SHTTP_X_FORWARDED_FOR เป็น CGI Environment Variable ที่บอก IP ของ Client ในกรณีที่ Client ใช้งานผ่าน Proxy Server
- 16) Session ID เป็นตัวเลขที่เป็น Unique ของแต่ละ User ที่เข้ามาใช้งาน โดยตัวเลขนี้จะมีเวลาหรืออายุการใช้งาน 20 นาที และหลังจาก 20 นาทีจะมีการ generate Session ID ค่าใหม่ให้โดยอัตโนมัติ หรือ ทุกๆ 20 นาที Session ID จะเปลี่ยนค่าไป สาเหตุที่เลือกใช้เวลา 20 นาที เพราะได้กำหนดสมมติฐานว่า User 1 คนใช้งานเว็บไซต์โดยเฉลี่ยต่อครั้งอยู่ที่ 20 นาที

ได้รับข้อมูลจึงเป็นผลทำให้ค่าสถิติที่ได้ไม่ครอบคลุมถึงปริมาณการเยี่ยมชมดังกล่าว แต่อย่างไรก็ตามโดยทั่วไปแล้ว บราวเซอร์ของผู้เยี่ยมชมจะ Enable JavaScript เพราะ การ Enable JavaScript เป็นค่า Default สำหรับ Browser ทั่วไป และ การใช้ JavaScript นั้นเป็นที่นิยมและมีการใช้งานอย่างแพร่หลายเพื่อพัฒนาเว็บไซต์

3. สถิติที่ทฤษฎีสามารถวัดได้

จากข้อมูลต่างๆ ในข้อ 2 ระบบทฤษฎีนำมาประมวลผลและวิเคราะห์อยู่ในรูปแบบข้อมูลสถิติการเยี่ยมชมเว็บไซต์ ซึ่งมีรายละเอียดของข้อมูลดังตารางที่ 2

เนื่องจากทฤษฎีนั้นอาศัยการทำงานของ JavaScript ดังนั้นในกรณีที่บราวเซอร์ของผู้เยี่ยมชมได้ Disable JavaScript แล้วทฤษฎีจะไม่สามารถ

ตารางที่ 2. แสดงความหมายของข้อมูลสถิติที่น่าสนใจในระบบบริการทฤษฎี

สถิติข้อมูลที่น่าสนใจ	ชื่อตัวแปรที่เกี่ยวข้อง	ความหมาย และ ประโยชน์
Unique IP	1. \$REMOTE_ADDR 2. SHTTP_X_FORWARDED_FOR 3. \$REMOTE_ADDR	ดูจาก IP ของผู้เรียกใช้งานหน้าเว็บนั้นจริงๆ โดยแบ่งได้เป็น 3 กรณีคือ 1. กลุ่มผู้เยี่ยมชมเว็บไซต์ที่ไม่ได้ทำการเชื่อมต่อผ่าน proxy server ในกรณีนี้ TrueHits จะทราบค่า IP ของผู้เยี่ยมชมเว็บไซต์จริงๆ 2. กลุ่มผู้เยี่ยมชมเว็บไซต์ที่ทำการเชื่อมต่อผ่าน proxy server โดยที่ proxy server ส่งผ่านค่า IP ของเครื่องผู้เยี่ยมชมเว็บไซต์ ซึ่งทำให้ TrueHits ทราบค่า IP ของผู้เยี่ยมชมเว็บไซต์จริงๆ 3. กลุ่มผู้เยี่ยมชมเว็บไซต์ที่ทำการเชื่อมต่อผ่าน proxy server โดยที่ proxy server ไม่ส่งผ่านค่า IP ของเครื่องผู้เยี่ยมชมเว็บไซต์ ในกลุ่มนี้ TrueHits จะทราบค่า IP ของ Proxy server เท่านั้น แต่ไม่สามารถทราบค่า IP ของผู้เยี่ยมชมเว็บไซต์จริงๆ ดังนั้น จึงใช้ค่า IP ของ Proxy Server แทน IP ของผู้เยี่ยมชมเว็บไซต์จริงๆ
Unique Session	Session ID	จำนวน Session ที่ไม่ซ้ำกัน ซึ่งใช้เป็นดัชนีวัดจำนวนผู้เยี่ยมชมเว็บไซต์ โดยใช้เทคนิคคุกกี้ (Cookie+Session) กล่าวคือ TrueHits จะส่งผ่าน cookie ไปติดตั้งไว้ที่ Browser ของผู้เยี่ยมชม และมีเลข ID Number ที่ถูก random ขึ้น ID Number นี้จะเป็นตัวอ้างอิงผู้เยี่ยมชมเว็บไซต์ โดยในปัจจุบัน TrueHits กำหนดอายุของ Session ไว้เป็น 20 นาที
PageViews	-	จำนวนครั้งที่เว็บเพจนั้นถูกเรียกจากผู้เยี่ยมชม ซึ่งจะเท่ากับค่าของ Counter ทั่วไปที่ใช้นับจำนวนฮิต
Return Visitor	Unique IP	จำนวนผู้เข้าชมที่เข้ามาดูแล้ว กลับมาดูเว็บไซต์นั้นอีก โดยดูจากค่าของ IP เดิม
Top Visitor	Unique IP	แสดง IP ที่เข้ามาดูเว็บไซต์โดยเรียงจากการใช้งานมากไปน้อย พร้อมกับบอกด้วย

สถิติข้อมูลที่น่าสนใจ	ชื่อตัวแปรที่เกี่ยวข้อง	ความหมาย และ ประโยชน์
		ว่ามาจาก ISP รายใด
Top ISP	Unique IP	แสดงกลุ่มของ ISP ของคนที่เข้ามาดูเว็บนั้น โดยเรียงจากมากไปน้อย
Top Referrer Domain	Referrer URL	แสดง Referrer Domain โดยเรียงจากมากไปน้อย
Top Os	SHTTP_USER_AGENT	แสดง OS ของผู้ที่เข้ามาดูเว็บไซต์โดยเรียงจากมากไปน้อย
Top Browser	SHTTP_USER_AGENT	แสดง Browser ของผู้ที่เข้ามาดูเว็บไซต์โดยเรียงจากมากไปน้อย
Top Search Engine	Referrer URL	แสดง Search Engine ที่ใช้ในการค้นหาเว็บไซต์ของท่านว่ามาจากที่ไหน
Top Search Keyword	Referrer URL	แสดง Search Keyword ที่ใช้ในการค้นหาเว็บไซต์ของท่าน
Top Language	SHTTP_ACCEPT_LANGUAGE	แสดงรายชื่อภาษาของผู้ที่เข้ามาดูเว็บไซต์ได้กำหนดไว้ที่ Web Browser
Top Page	Page Name	ในกรณีที่ สมาชิกระบุชื่อ page ที่ใช้ในการตรวจสอบสถิตินั้น ระบบจะแสดง รายชื่อPage ที่ได้รับรู้ว่าหน้าใดที่มีการเข้ามาเรียกดูมากที่สุด
Top Proxy Server	SHTTP_VIA	ในกรณีที่ user เรียกเว็บไซต์ผ่านทาง Proxy Server จะแสดงข้อมูลของ Proxy Server โดยบอกรายละเอียดเพิ่มเติมด้วยว่า proxy server แต่ละตัวปล่อย IP Client หรือไม่
Top Resolution, Top Color Depth	Resolution, Color Depth	แสดงความละเอียดหน้าจอและการเซตบิตสีของ web browser ของผู้ที่เข้ามาเยี่ยมชมเว็บไซต์
Enable Setting	Java Enable, Cookie Enable และ JavaScript Version	ตรวจสอบว่า Web Browser ของผู้เยี่ยมชม Enable Java หรือ Enable Cookie ไว้ที่เปอร์เซ็นต์ และใช้ JavaScript Version ใดมากที่สุด

จากความหมายของสถิติข้อมูล ระบบทรูฮิตใช้ Unique IP แทนจำนวนผู้เยี่ยมชมที่ไม่ซ้ำกัน ส่วนระบบตรวจสอบสถิติของต่างประเทศที่ได้รับความนิยมต่างก็นิยมการวัดจำนวนผู้เยี่ยมชมแตกต่างกันออกไป ดังรายละเอียดต่อไปนี้ คือ

- Hit Box ใช้ Session ID ที่มีอายุ 24 ชั่วโมง แทนจำนวนผู้เยี่ยมชม
- Extreme ใช้ Session ID ที่มีอายุ 15 นาที แทนจำนวนผู้เยี่ยมชม
- Stat4All ใช้ Session ID ที่มีอายุ 20 นาที แทนจำนวนผู้เยี่ยมชม
- Ned Stat แสดงจำนวน pageviews หรือจำนวน click เท่านั้น
- ทรูฮิต ใช้ Unique IP เป็นตัววัดจำนวนผู้เยี่ยมชม สามารถใช้ IP เป็นตัว track user ได้ในกรณีข้ามวัน เช่น กลับเข้ามาใช้งานซ้ำกันเท่าไรในช่วงเวลา 1 เดือน ซึ่งในกรณีที่ใช้ Session ID ที่มีอายุ 24 ชั่วโมง หรือ 15 นาที สามารถ track user ได้ภายในช่วงเวลา 24 ชั่วโมง และ 15 นาที ได้เท่านั้นตามลำดับ

สาเหตุที่ทรูฮิตใช้ตัววัด 3 ตัวในการวัดปริมาณผู้เยี่ยมชมเว็บไซต์ ได้แก่ Unique Visitor, Unique Session และ Pageview โดยที่ทรูฮิตใช้ Unique

Visitor ซึ่งคำนวณจากจำนวน IP ของผู้เยี่ยมชมที่ไม่ซ้ำกัน โดยที่จุดเด่นอยู่ที่ IP เป็นตัววัดที่ให้ความเชื่อถือได้สูงสุดเมื่อเทียบกับ Unique Session และ Pageview เพราะปลอมแปลงได้ยากที่สุด แต่มีข้อเสียที่การคำนวณหาขึ้นต้องอาศัย Computation Power สูงกว่าตัวอื่นมาก เนื่องจากต้องเก็บค่า IP ที่ไม่ซ้ำกันในเวลานานถึงเดือนเข้าไปใน Memory เพื่อคำนวณค่า IP ที่ไม่ซ้ำกันได้ ซึ่งแตกต่างจากการใช้ Unique Session ซึ่งเก็บค่า Session ที่ไม่ซ้ำกันตามช่วงเวลาที่กำหนดไว้เท่านั้น คือ 20 นาที ซึ่งค่าที่นำมาคำนวณไม่มาก ทำให้ไม่เสียเวลาและ memory ในการคำนวณมากเท่ากับการทำ Unique IP ดังกล่าว

ดังจะเห็นได้ว่าผู้ให้บริการที่มีชื่อเสียงในต่างประเทศนั้นจะไม่นำเสนอข้อมูลของ Unique Visitor โดยคำนวณจาก IP ส่วนมากจะเสนอในรูปแบบของ Unique Session โดยใช้ Session ID ที่มีเวลาต่างกันออกไป

ในส่วน Unique Session ที่ทรูฮิตใช้ Session ID 20 นาทีนั้น ในปัจจุบันยังไม่มีมาตรฐานที่ชัดเจน แต่อย่างไรก็ตามค่า Unique Session (Session ID 20 นาที) ซึ่งหมายถึง อายุของ Session ในการใช้งานเว็บไซต์ของ

เครื่องคอมพิวเตอร์ 1 เครื่องมีเวลาการใช้งานอยู่ที่ 20 นาที ถ้าหมดอายุแล้วระบบจะให้ค่า Session ID ใหม่ทันที

โดยระบบทฤษฎีนำเสนอตัวแปรทั้ง 3 ตัวเพื่อความถูกต้องในการวิเคราะห์จำนวนผู้เยี่ยมชมเว็บไซต์ได้ดีกว่าการใช้อ้างอิงด้วยค่าใดค่าหนึ่งเท่านั้น โดยแต่ละตัวแปร มีความสัมพันธ์ดังนี้

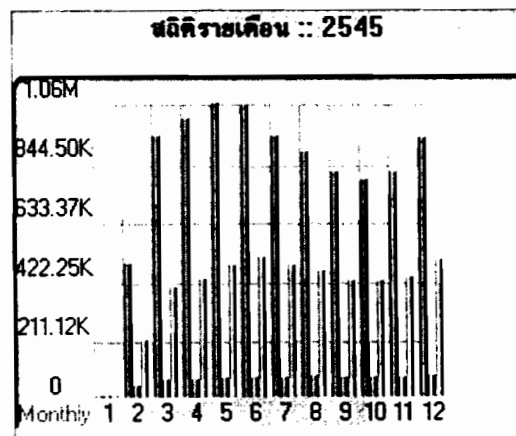
Unique Visitor (IP) น้อยกว่าเท่ากับ Unique Session น้อยกว่าเท่ากับ Pageview

หรือ Daily Visitor (IP 1 วัน) มีค่าประมาณ Unique Session ที่มี Session time เป็น 1 วัน

หรือ Page view มีค่าประมาณ Unique Session ที่มี Session time เป็น 0

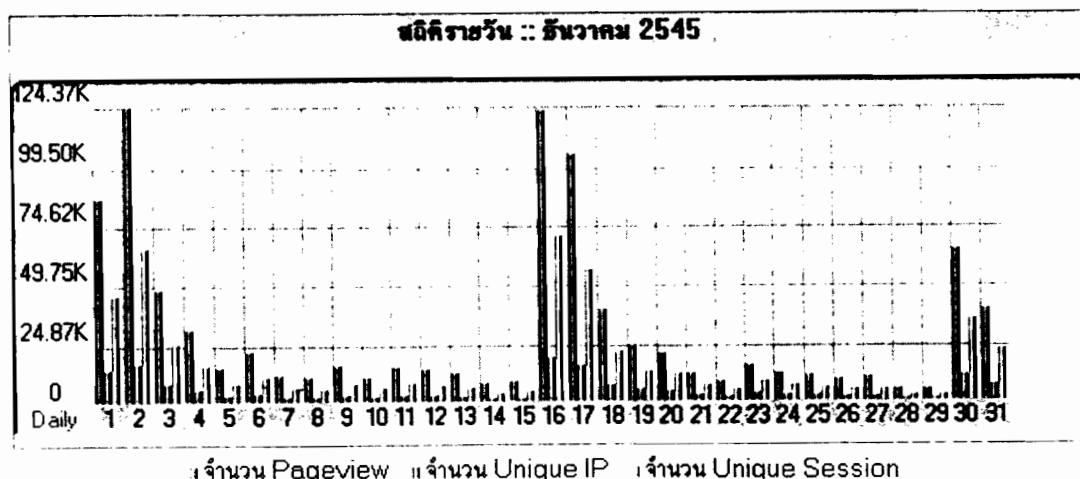
จากตารางที่ 2 ที่อธิบายความหมายของสถิติข้อมูล ในบทความนี้ขอยกตัวอย่างการใช้งานจริงของเว็บสำนักงานสลาकिनแบ่งรัฐบาลเพื่อให้เห็นถึงวิธีการวิเคราะห์สถิติที่จะเป็นแนวทางให้เกิดประโยชน์ได้ ดังตัวอย่างต่อไปนี้

3.1) ตัวอย่างที่นำเสนอในส่วนของ ตัวแปร Unique IP, Unique Session, Pageviews มีทั้งข้อมูลรายเดือน ข้อมูลรายวัน และข้อมูลรายชั่วโมง ดังรูปที่ 4, 5 และ 6 ตามลำดับ



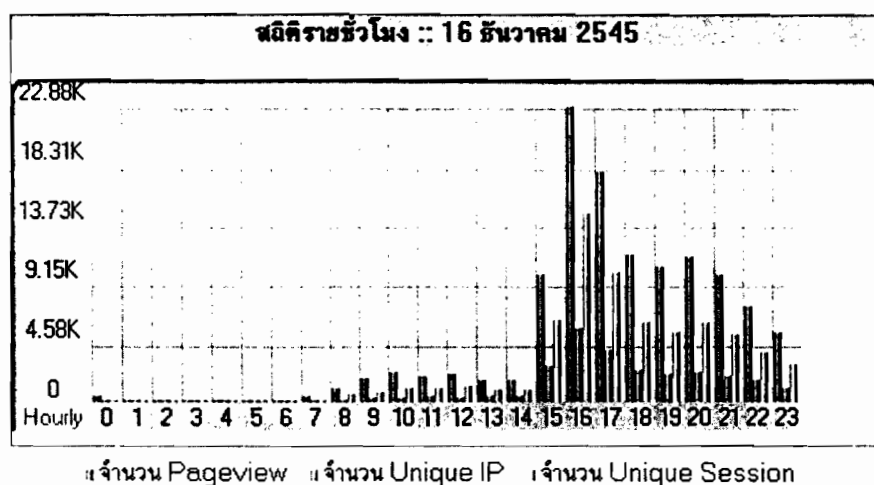
จำนวน Pageview จำนวน Unique IP จำนวน Unique Session

รูปที่ 4. แสดงข้อมูลสถิติของ Pageviews, Unique IP, Unique Session รายเดือน



จำนวน Pageview จำนวน Unique IP จำนวน Unique Session

รูปที่ 5. แสดงข้อมูลสถิติของ Pageviews, Unique IP, Unique Session รายวัน



รูปที่ 6. แสดงข้อมูลสถิติของ Pageviews, Unique IP, Unique Session รายชั่วโมง

จากรูปที่ 5 และ 6 จะเห็นว่า จำนวน Unique IP, Unique Session, Pageviews นั้นจะมีปริมาณสูงมากในวันที่ 1 และ 16 ของรายเดือน นอกจากนี้ จะมีปริมาณสูงมากในชั่วโมงตั้งแต่ 15 นาฬิกา เป็นต้นไป เนื่อง

จากเว็บไซต์นี้เป็นของสำนักงานสลากกินแบ่งรัฐบาล ซึ่งจะรายงานผลของสลากกินแบ่งรัฐบาล ทุกๆ วันที่ 1 และ 16 ของทุกเดือน หลังเวลา 15 นาฬิกา จึงเป็นผลทำให้สถิติพุ่งสูงขึ้นในวัน และเวลา ดังกล่าว เป็นต้น

3.2) ตัวอย่างที่นำเสนอในส่วนของตัวแปร Top Visitor ดังรูปที่ 7

TOP 30 Daily Visitor		
IP	From ISP	Pageview
203.113.86.130	TOT-TH	93
172.20.8.31	Private IP	89
202.47.247.146	CAT9931 route object	67
203.149.57.10	Corporate Networks of Samart Infonet Co.,Ltd.	64
203.146.147.13	CAT route object for LoxInfo	54
210.4.136.251	INET-TH	52
202.247.13.240	BIGLOBE CIDR BLOCK 1	51

รูปที่ 7. แสดงข้อมูล Top Visitor

จากรูปที่ 7 ทำให้เห็นว่า IP ใดที่เข้ามาใช้งานมากที่สุด และมาจาก ISP ใด และมีปริมาณการใช้งานเท่าไร

3.3) ตัวอย่างที่นำเสนอในส่วนของตัวแปรของ Top ISP ดังรูปที่ 8

TOP ISP			
NAME	Number		Percent
TOT-TH	1031	=====	11.89%
KSC Commercial Internet	731	=====	8.43%
INET-TH	716	=====	8.26%
CAT route object for LoxInfo	531	=====	6.12%
UNINET (THAINET) block	290	=====	3.34%
CAT route object for C.S.Internet	284	=====	3.28%
CAT route object for CS Communication	281	=====	3.24%
AsiaNet	271	=====	3.13%

รูปที่ 8. แสดงข้อมูล Top ISP

จากรูปที่ 8 ทำให้เห็นว่า ISP ใดที่เข้ามาใช้งานมากที่สุด และมีปริมาณการใช้งานเท่าไร

3.4) ตัวอย่างที่นำเสนอในส่วนของตัวแปรของ Top OS ดังรูปที่ 9

TOP OS			
NAME	Number		Percent
Windows 98	6590		76.49%
Windows NT 5.0(Windows 2000)	870		10.10%
Windows NT 5.1(Windows XP)	785		9.11%
Windows 95	197		2.29%
Windows NT 4.0	102		1.18%
Windows NT	60		0.70%
Linux 2.2.16 i686	6		0.07%
Linux i586	2		0.02%

รูปที่ 9. แสดงข้อมูล Top OS

จากรูปที่ 9 ทำให้เห็นว่า OS ใดเป็นที่นิยมใช้งานมากที่สุด และมีปริมาณการใช้งานเท่าไร

3.5) ตัวอย่างที่นำเสนอในส่วนของตัวแปร Top Browser ดังรูปที่ 10

TOP Browser Version			
NAME	Number		Percent
MSIE 5.0	2698		32.63%
MSIE 6.0	2295		27.75%
MSIE 5.5	2090		25.28%
MSIE 5.01	662		8.01%
MSIE 6.0b	345		4.17%
MSIE 4.01	166		2.01%
Mozilla ux 2.2.16 i686	6		0.07%

รูปที่ 10. แสดงข้อมูล Top Browser

จากรูปที่ 10 ทำให้เห็นว่า Browser ใดเป็นที่นิยมใช้งานมากที่สุด และมีปริมาณการใช้งานเท่าไร

3.6) ตัวอย่างที่นำเสนอในส่วนของตัวแปร Top Resolution ดังรูปที่ 11

Resolution			
NAME	Number		Percent
800*600	74919		75.40%
1024*768	21185		21.32%
640*480	1496		1.51%
1152*864	788		0.79%

รูปที่ 11. แสดงข้อมูล Top Resolution

จากรูปที่ 11 ทำให้เห็นว่า Resolution ใดเป็นที่นิยมมากที่สุด และมีปริมาณการใช้งานเท่าไร

3.7) ตัวอย่างที่นำเสนอในส่วนของตัวแปร Top Search Engine และ Search Keyword ดังรูปที่ 12

TOP Search Engine			
NAME	Number		Percent
google.com	522		57.14%
yahoo.com	361		39.56%
altavista.com	18		1.10%
siamguru.com	16		1.10%
aol.com	11		1.10%

TOP Search Keyword			
NAME	Number		Percent
thailand picture	188		34.08%
picture of thailand	161		29.16%
การพัฒนาบุคลากร	96		17.39%
งานวิจัย	68		12.31%

รูปที่ 12. แสดงข้อมูล Top Search Engine และ Search Keyword

จากรูปที่ 12 ทำให้เห็นว่า Search Engine และ Search Keyword ใดเป็นที่นิยมมากที่สุด และมีปริมาณการใช้งานเท่าไร

จากตัวอย่างดังกล่าวที่นำเสนอ นักพัฒนาเว็บไซต์สามารถใช้เป็นข้อมูลเพื่อช่วยในการปรับปรุงและพัฒนาเว็บไซต์ให้ดียิ่งขึ้น และนอกจากนี้สามารถนำข้อมูลมารวบรวมและวิเคราะห์สถิติภาพรวมได้ดังหัวข้อต่อไป

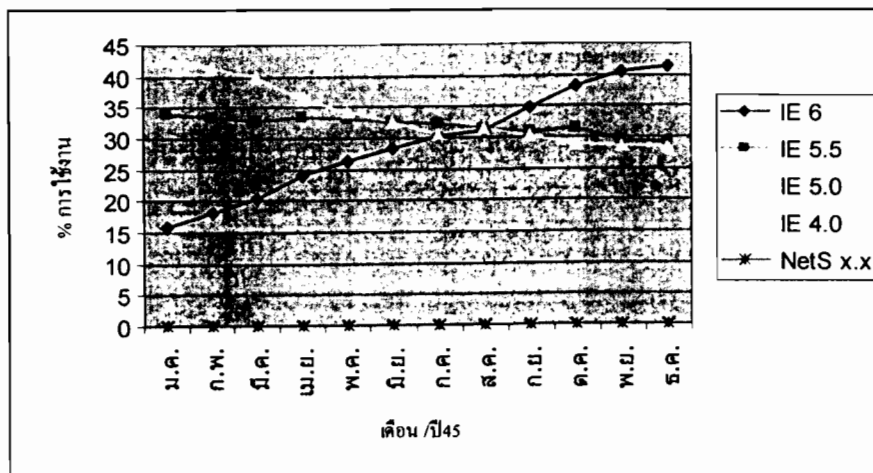
4. สถิติภาพรวมที่ได้จากระบบทฤษฎี

ระบบทฤษฎีทำการประมวลผลวิเคราะห์ข้อมูลการเยี่ยมชมเว็บไซต์ของเว็บไซต์สมาชิกทั้งหมดในช่วงเวลาตั้งแต่ 1 มกราคม 2545 ถึง 31 ธันวาคม 2545 ซึ่งมีจำนวนข้อมูล 2,354,012,488 ชุดข้อมูล (Record) จากการประมวลผลชุดข้อมูลดิบทั้งหมดเพื่อคำนวณหาค่า Average Monthly Unique IP พบว่ามีจำนวนถึง 2,392,504 ซึ่งเป็นตัวเลขที่แสดงถึงจำนวน

เฉลี่ยของผู้เยี่ยมชมที่ไม่ซ้ำกัน (Unique Visitors) ใน 1 เดือนหรืออีกนัยหนึ่งแสดงถึงจำนวนผู้ใช้อินเทอร์เน็ต มีจำนวนขั้นต่ำในช่วงหนึ่งเดือน โดยที่สถิติภาพรวมการเยี่ยมชมเว็บไซต์ของสมาชิกทั้งหมด สะท้อนให้เห็นถึงพฤติกรรมการใช้อินเทอร์เน็ตในประเทศนั้นเป็นอย่างไร รวมถึงแนวโน้มของการใช้ OS หรือ Browser เป็นอย่างไรด้วย ยกตัวอย่างเช่น

4.1 สถิติการใช้ Web Browser โดยนำเสนอเป็นรายเดือนใน 1 ปี และแกน Y เป็น % การใช้งาน

ค่าของ Browser ได้ข้อมูลจากตัวแปร \$HTTP_USER_AGENT จาก CGI Environment variable

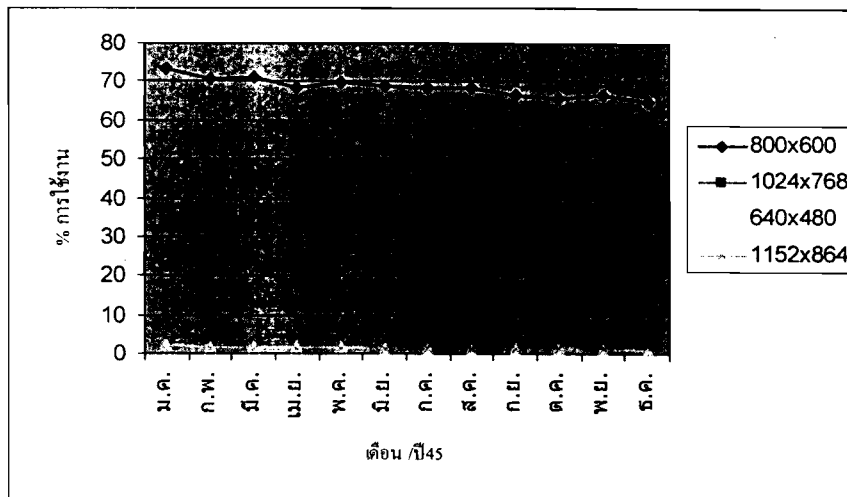


รูปที่ 13. แสดงข้อมูลสถิติการใช้อินเทอร์เน็ตโดยดูจาก Web Browser

จากรูปที่ 13 ทำให้เห็นได้ว่า IE 6 มีการใช้งานเพิ่มขึ้นทุกๆ เดือน ส่วน IE 5.5, IE 5.0, และ IE4.0 ต่างมีแนวโน้มใช้งานลดลง และ Netscape นั้นมีการใช้งานที่น้อยมาก และตั้งแต่เดือนกรกฎาคม พ.ศ.2545 IE 6 เริ่มมีอัตราการใช้งานแพร่หลายเหนือกว่า IE 5

4.2 สถิติความละเอียดของหน้าจอ โดยนำเสนอเป็นรายเดือนใน 1 ปี และแกน Y เป็น % การใช้งาน

ค่าของความละเอียดของหน้าจอได้ข้อมูลจากตัวแปร screen.width+'*'+screen.height จาก โค้ด JavaScript

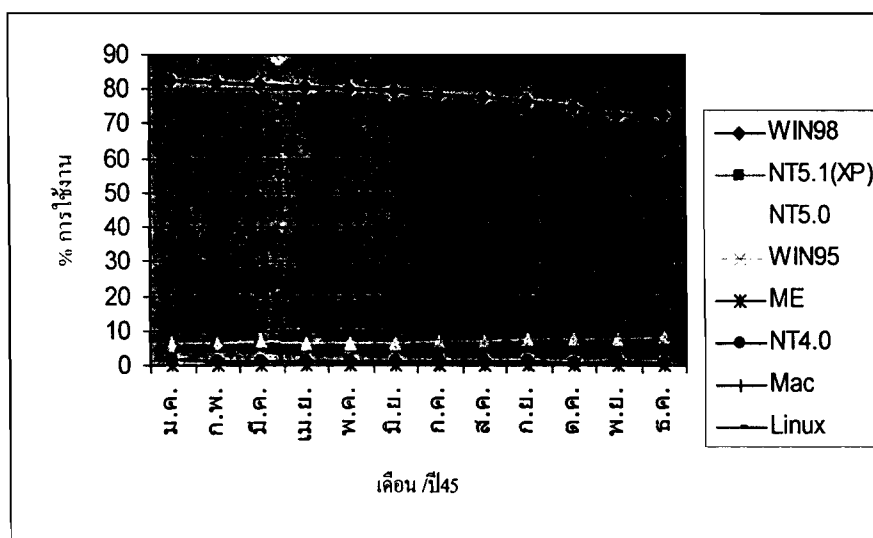


รูปที่ 14. แสดงข้อมูลสถิติการใช้อินเทอร์เน็ต โดยดูจาก ความละเอียดของหน้าจอ

จากรูปที่ 14 ทำให้เห็นได้ว่า ความละเอียดของหน้าจอ 800*600 ก็ยังคงเป็นที่นิยมมากที่สุด และ ความละเอียดของหน้าจอ 1024*768 ก็มีแนวโน้มที่เพิ่มขึ้นตลอด ดังนั้นการออกแบบเว็บไซต์ควรจะออกแบบโดยเน้นให้รองรับหน้าจอความละเอียด 800 x 600 เป็นหลัก เพราะมีผู้ใช้อินเทอร์เน็ตเฉลี่ยกว่า 60 % ใช้ความละเอียดหน้าจอดังกล่าว

4.3) สถิติ ระบบปฏิบัติการ (OS) โดยนำเสนอเป็นรายเดือนใน 1 ปี และแกน Y เป็น % การใช้งาน

ค่าของระบบปฏิบัติการ ได้ข้อมูลจากตัวแปร SHTTP_USER_AGENT จาก CGI Environment variable

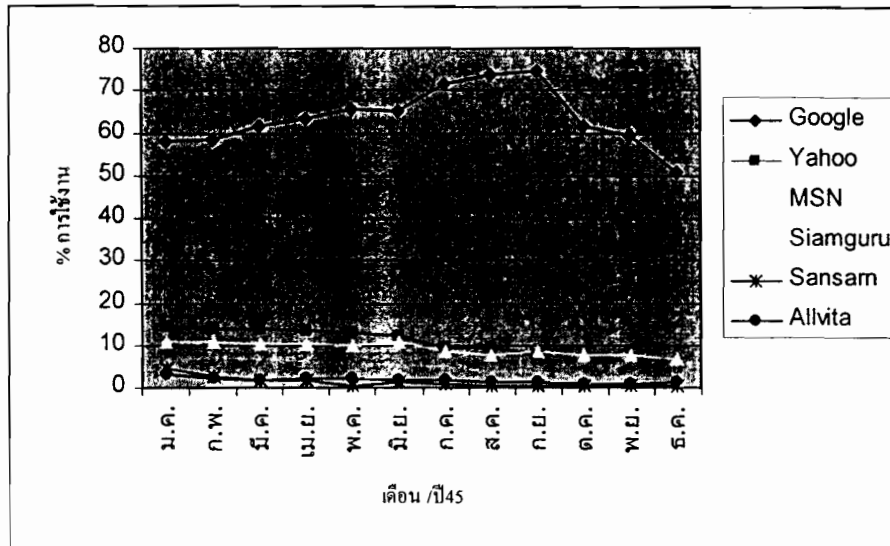


รูปที่ 15. แสดงข้อมูลสถิติการใช้อินเทอร์เน็ต โดยดูจาก ระบบปฏิบัติการ

จากรูปที่ 15 ทำให้เห็นได้ว่า ระบบปฏิบัติการ Window 98 ยังคงได้รับความนิยมอยู่มาก แม้ว่า Window XP จะมีแนวโน้มการใช้งานที่เพิ่มมากขึ้นก็ตาม แต่ยังคงน้อยกว่าและห่างจาก Window 98 อยู่ค่อนข้างมาก และที่สำคัญระบบปฏิบัติการ Linux สำหรับเครื่อง Desktop Computer นั้น ยังไม่ได้รับความนิยมเท่าที่ควร โดยมีอัตราการใช้งานเฉลี่ยราว 0.07%

4.4 สถิติการใช้ Search Engine โดยนำเสนอเป็นรายเดือนใน 1 ปี และแกน Y เป็น % การใช้งาน

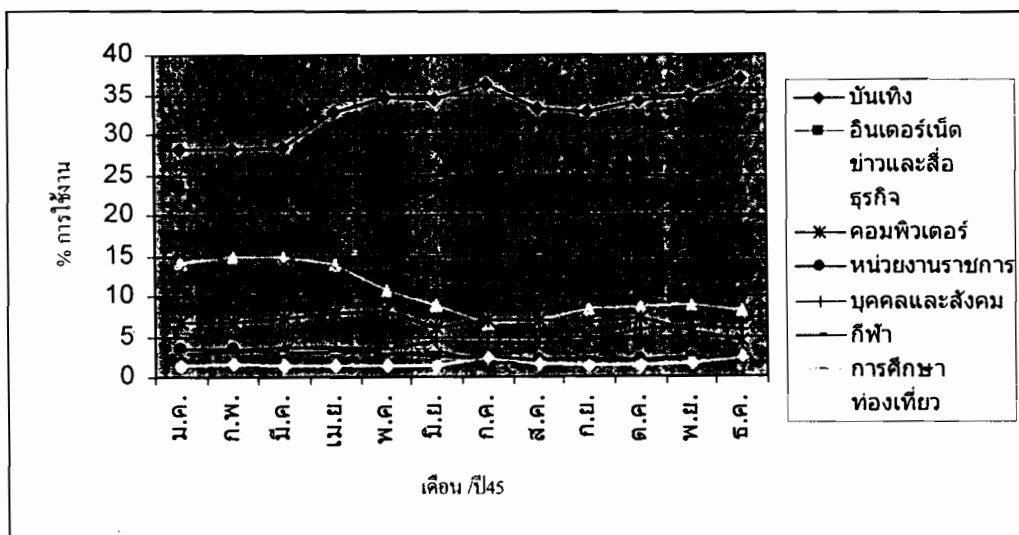
ค่าของ Search Engine URL ได้ข้อมูลจากตัวแปร `escape(document.referrer)` จาก โค้ด JavaScript



รูปที่ 16. แสดงข้อมูลสถิติการใช้อินเทอร์เน็ตโดยดูจากการใช้ search engine

จากรูปที่ 16 ทำให้เห็นได้ว่า google.com เป็น search engine ที่ได้รับความนิยมสูงสุด และเริ่มมีอัตราการใช้งานลดลงตั้งแต่ช่วงเดือนกันยายน ในขณะที่ yahoo.com ซึ่งได้รับความนิยมเป็นอันดับสอง นั้นเริ่มมีอัตราการใช้งานเพิ่มขึ้นอย่างเห็นได้ชัดในช่วงเวลาดังกล่าว

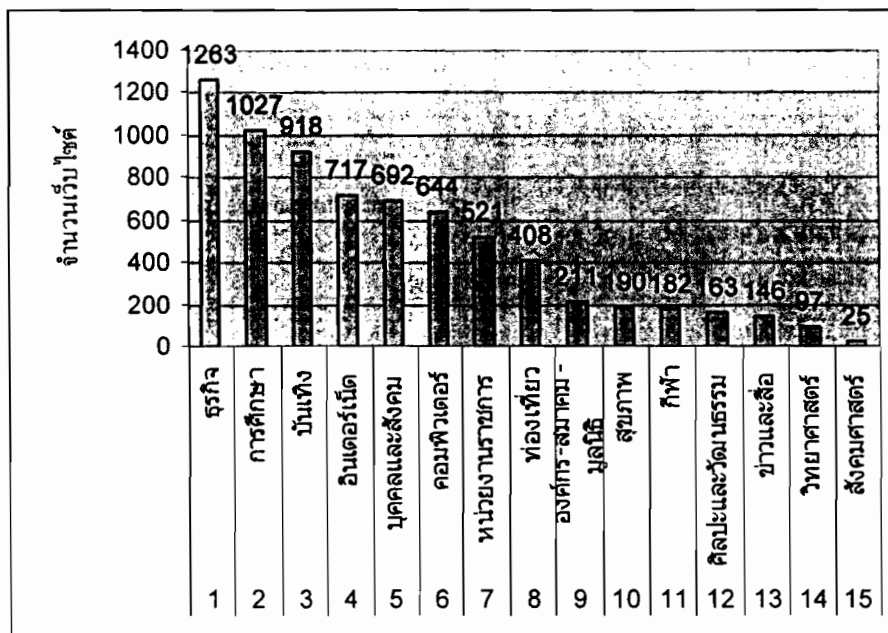
4.5 สถิติการใช้อินเทอร์เน็ตของคนไทย นิยมใช้อินเทอร์เน็ตเพื่ออะไรมากที่สุด



รูปที่ 17. แสดงข้อมูลสถิติการใช้อินเทอร์เน็ตเพื่ออะไร

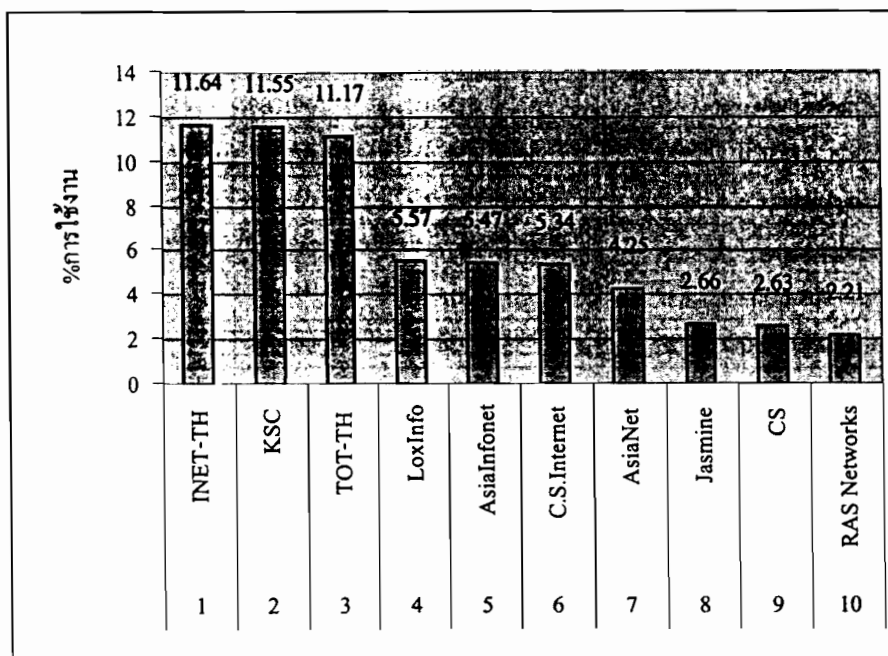
จากรูปที่ 17 จะเห็นได้ชัดเจนว่า คนไทยนิยมใช้อินเทอร์เน็ตเพื่อความบันเทิง อินเทอร์เน็ต ข่าวและสื่อ เป็นส่วนใหญ่ และมีแนวโน้มที่เพิ่มขึ้นเรื่อยๆ ในขณะที่ ส่วนของเว็บไซต์ของหน่วยงานราชการ และเว็บไซต์

ทางการศึกษาในหมวดโรงเรียน มหาวิทยาลัยนั้นก็ไม่ค่อยได้รับความนิยม ทั้งๆ ที่จำนวนเว็บไซต์ในหมวดดังกล่าวมีปริมาณมากพอสมควร ดังแสดงในกราฟรูปที่ 18



รูปที่ 18. แสดงจำนวนสมาชิกที่เข้ามาใช้งานทูลูติคโดยจัดแบ่งตามแต่ละหมวด โดยที่แกน Y เป็นจำนวนเว็บไซต์

4.6 สถิติการการจัดอันดับ Top ISP 10 อันดับแรก ที่ได้จากคนเล่นเว็บไซต์ในเมืองไทย



รูปที่ 19. Top ISP 10 อันดับแรก เพื่อดูว่า ในประเทศไทย ISP รายไหนที่ได้รับความนิยมเรียงจากมากไปน้อย

จากตัวอย่างในรูปที่ 13-19 ทำให้เห็นถึงภาพรวมการใช้งานอินเทอร์เน็ตของประเทศไทยได้ ยกตัวอย่างเช่น คนใช้อินเทอร์เน็ตเพื่ออะไร ซึ่งแสดงให้เห็นถึงพฤติกรรมการใช้งานของกลุ่มคนส่วนใหญ่ได้ เป็นต้น

5. จัดอันดับเว็บไซต์ประจำปี 2545

ระบบทูลิตได้จัดอันดับเว็บไซต์ประจำปี 2545 จากค่าโดยเฉลี่ยของ Unique IP รายเดือนในปี 2545 เพื่อแสดงให้เห็นว่าเว็บไซต์ใดเป็นที่นิยมมากที่สุดในเมืองไทย โดยแสดง 20 อันดับเว็บไซต์แรกดังตารางที่ 3 และแสดง 20 อันดับเว็บไซต์ของหน่วยงานราชการ ดังตารางที่ 4 นอกจากนี้ระบบได้นำเสนอตัวแปรดัชนีการจัดเรียงอีกอย่างหนึ่ง คือ ค่าความนิยม (Popularity Rate) โดยมีสูตรคือ

$$\text{ความนิยม (Popularity)} = 100 \times \text{Unique IP ของเว็บไซต์} / \text{Unique IP ของเว็บไซต์สมาชิกทั้งหมด}$$

ซึ่งค่าความนิยม (Popularity) เป็นค่าที่แสดงถึงอัตราความนิยมของเว็บไซต์ได้ชัดเจนกว่าการวัดโดยใช้ปริมาณ Unique IP อย่างเดียว ดังตัวอย่างเช่น ในกรณีที่เว็บไซต์หนึ่งๆ จะมีการเจริญเติบโตของจำนวน Unique IP ขึ้นเรื่อยๆ นั้น มิได้แสดงว่า เว็บไซต์นั้นได้รับความนิยมเพิ่มขึ้นเสมอไป เพราะถ้าเว็บไซต์นั้นมีค่า Popularity Rate ลดลง แสดงได้ว่าอัตราการเยี่ยมชมลดลงเมื่อเทียบกับจำนวนผู้เยี่ยมชมเว็บไซต์ทั้งหมด

ตารางที่ 3. แสดง Top 20 Website จากสมาชิกประมาณ 7000 เว็บไซต์ ประจำปี 2545

ลำดับ	ชื่อเว็บไซต์	IP เฉลี่ยรายเดือน	Session เฉลี่ยรายเดือน	Pageviews เฉลี่ยรายเดือน	Popularity Rate (%)
1	www.sanook.com	313554	15064686	25640860	16.74
2	www.esmszone.com	215186	1216484	1678161	11.49
3	www.thaimisc.com	212383	4426417	8730232	11.34
4	www.mthai.com	200492	5307048	9363385	10.70
5	www.dailynews.co.th	191223	765076	958975	10.21
6	mweb.co.th	180381	3690289	6690956	9.63
7	www.thaiware.com	157855	1873435	2942825	8.43
8	www.matichon.co.th	151429	1581106	3225272	8.08
9	www.siamsport.co.th	145941	2828733	5952738	7.79
10	www.komchadluek.com	135801	591089	731521	7.25
11	www.songtoday.com	134701	1628992	2544439	7.19
12	www.bangkokbiznews.com	110224	708842	1237864	5.88
13	www.madoo.com	103785	2016923	4013397	5.54
14	www.ragnarok.in.th	101856	1729556	2582439	5.44
15	www.keepalbum.com	98674	990936	1931369	5.27
16	www.ThaiSecondhand.com	95628	2016028	4248640	5.10
17	www.manager.co.th	92379	847977	1590545	4.93
18	www.pramool.com	79223	2723025	5483803	4.23
19	www.mcot.or.th	70867	436455	684143	3.78
20	www.teenec.com	55806	838780	1322630	2.98

จากตารางที่ 3 จะเห็นได้ชัดเจนว่า เว็บไซต์อยู่ในอันดับ 20 เว็บไซต์ยอดนิยม นั้น จะเป็นเว็บทางด้านบันเทิงเป็นส่วนใหญ่ และ เว็บไซต์ได้อันดับ 1 นั้นมี Unique IP สูงกว่าเว็บอื่นใน 20 อันดับแรกมาก และในกรณีของ

sanook ที่มีค่า popularity=16.74 หมายความว่าถ้ามีกลุ่มคนเล่นเว็บไซต์จำนวน 100 คน จะมีคนเข้ามาเยี่ยมชมเว็บของ sanook เฉลี่ยถึง 16.74 คน เป็นต้น

ตารางที่ 4. แสดง Top 20 website ของหน่วยงานราชการ ประจำปี 2545

ลำดับ	ชื่อเว็บไซต์	IP เฉลี่ยรายเดือน	Session เฉลี่ยรายเดือน	Pageviews เฉลี่ยรายเดือน	Popularity Rate (%)
1	www.tat.or.th	90525	201763	267606	4.83
2	www.glo.or.th	71859	428088	874990	3.83
3	www.mcot.or.th	70867	436455	684143	3.78
4	truchits.net	41531	147993	211436	2.21
5	www.bot.or.th	32973	120957	164374	1.76
6	www.nectec.or.th	31369	76163	90449	1.67
7	www.dola.go.th	27851	88563	127123	1.48
8	www.school.net.th	26032	150643	207306	1.39
9	www.moe.go.th	24290	100859	133581	1.29
10	www.police.go.th	22839	54254	67642	1.21
11	www.suparsit.com	21486	98555	186397	1.14
12	www.thairegistration.com	21315	63477	70774	1.13
13	www.mfa.go.th	20153	105241	187947	1.07
14	www.moc.go.th	17005	53147	64274	0.90
15	www.moi.go.th	16297	54379	99556	0.87
16	www.thaigov.go.th	16283	46966	52833	0.869
17	www.ge.go.th	16157	65126	80767	0.863
18	www.prd.go.th	11824	29323	39751	0.63
19	thaicert.nectec.or.th	9894	25539	29409	0.52
20	www.rd.go.th	7501	26391	29273	0.40

จากตารางที่ 4 แสดง Top 20 เว็บไซต์ของหน่วยงานราชการนั้น Unique IP อยู่ในระดับหลักหมื่น และแตกต่างจากเว็บในหมวดบันเทิงค่อนข้างมาก จะเห็นได้ว่าเว็บไซต์ต่างๆ ที่เป็นสมาชิกธุรกิจ สามารถนำมาจัดอันดับเว็บไซต์ได้ เนื่องจากใช้ตัววัดที่มีมาตรฐานเดียวกัน ทำให้สมาชิกู้อันดับและความนิยมของเว็บไซต์ตัวเอง และสามารถเปรียบเทียบกับเว็บในกลุ่มเดียวกันได้

6. Conclusion

ระบบให้บริการธุรกิจเป็นระบบบริการตรวจสอบและประมวลผลสถิติการเยี่ยมชมเว็บไซต์ของแต่ละเจ้าของเว็บไซต์โดยที่สามารถเข้ามาสมัครใช้บริการได้ และสามารถตรวจสอบสถิติต่างๆ เพื่อใช้เป็นแนวทางในการพัฒนาเว็บไซต์ให้มีคุณภาพมากยิ่งขึ้น และพัฒนาเนื้อหาในเว็บไซต์ให้ตรงกับกลุ่มคนที่เข้ามาเยี่ยมชมเว็บไซต์ได้ โดยในปัจจุบันระบบธุรกิจ

สามารถรองรับปริมาณกราฟฟิกันได้จำนวนสูงไม่น้อยกว่า 15 ล้าน transactions/วัน

ยิ่งไปกว่านั้นระบบธุรกิจ นำเสนอสถิติภาพรวมการใช้ Web browser, ความละเอียดของหน้าจอ, ระบบปฏิบัติการ, Search Engine ตลอดจนความนิยมในการใช้อินเทอร์เน็ตแบ่งตามหมวดหมู่ซึ่งได้จากการประมวลผลจากข้อมูลการใช้งานจริงจำนวน 2,354,012,488 ชุดข้อมูล ในระหว่างวันที่ 1 มกราคม ถึง 31 มกราคม พ.ศ. 2545 จากการประมวลผลชุดข้อมูลดิบทั้งหมดเพื่อคำนวณค่า Average Monthly Unique IP พบว่ามีจำนวนถึง 2,392,504 Unique IP ซึ่งเป็นตัวเลขที่แสดงถึงจำนวนเฉลี่ยของผู้เยี่ยมชมที่ไม่ซ้ำกัน (Unique Visitors) มีจำนวนขึ้นต่ำในช่วงหนึ่งเดือน

และที่สำคัญสามารถใช้ธุรกิจเป็นตัววัดโดยมีมาตรฐานเดียวกัน ซึ่งสามารถนำข้อมูลดังกล่าวมาเปรียบเทียบกันได้ เป็นผลให้มีการจัด Ranking เพื่อใช้เปรียบเทียบกับเว็บไซต์อื่นที่อยู่ในกลุ่มเดียวกัน

7. Reference

- [1] "Fast Counter", <http://www.bcentral.com/products/fc/default.asp>
- [2] Tec-Ed, Inc., "Accessing Web Site Usability from Server Log Files", White Paper <http://www.teced.com/PDFs/whitepap.pdf>, December 1999
- [3] WebsideStory, "HitBox Web Analytics Technology :The Future of Internet Intelligence", WhitePaper, http://www.websidestory.com/_data/resources/whitepapers/hitboxtechnology2.pdf, 2000
- [4] "Hit Box", <http://www.hitbox.com>.
- [5] "Ned Stat", <http://www.nedstatbasic.net>.
- [6] "Extreme", <http://www.extreme-dm.com/tracking/?home>.
- [7] "Stats4All", <http://www.stats4all.com/asp/login.asp>.
- [8] "NITC", <http://www.nitc.go.th/internetuser/survey2002.html>.
- [9] Ralf Rantza and Holger Schwarz, "Multi-Tier Architecture for High-Performance Data Mining", <http://citeseer.nj.nec.com/rantza99multitier.html>.
- [10] "BZip2", <http://freeware.sgi.com/cd-1/relnotes/bzip2.html>
- [11] R.Fielding, J. Gettys, J. Mogul, H. Frystyk, "Hypertext Transfer Protocol -- HTTP/1.1", <http://www.ietf.org/rfc/rfc2616.txt>, June 1999.
- [12] "Environment Variables in Apache", <http://httpd.apache.org/docs/env.html>.



จารุวรรณ ละอองมลาย สำเร็จการศึกษาวិชากรรมศาสตรบัณฑิตภาควิชาอิเล็กทรอนิกส์ เกียรตินิยมอันดับ 1 จากสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (พ.ศ.2536-2540) และ

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า จากจุฬาลงกรณ์มหาวิทยาลัย (พ.ศ.2540-2542) ปัจจุบันเป็นพนักงานโครงการของสำนักบริการเทคโนโลยีสารสนเทศภาครัฐ มีความสนใจเป็นพิเศษด้าน Security



ลิทธิชัย สิลานันท์นูล สำเร็จการศึกษาวิชากรรมศาสตรบัณฑิตและวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า จากจุฬาลงกรณ์มหาวิทยาลัย เมื่อปี 2540 และ 2543 ตามลำดับ ปัจจุบันเป็น

พนักงานโครงการของสำนักบริการเทคโนโลยีสารสนเทศภาครัฐ มีความสนใจเป็นพิเศษด้าน System Administration และ Network Security



ปิยะ คัชชาวณิชัย สำเร็จการศึกษาวิชากรรมศาสตรบัณฑิตภาควิชาอิเล็กทรอนิกส์ เกียรตินิยมอันดับ 2 จากสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (พ.ศ.2529-2533) และ วิทยาการสารสนเทศมหาบัณฑิต (Master Degree of Information Sciences) เกียรตินิยมอันดับ 1 มหาวิทยาลัยโคโซกุ ประเทศญี่ปุ่น (พ.ศ. 2536-2538) และ วิทยาการสารสนเทศศษญีบัณฑิต (PhD. in Information Sciences) มหาวิทยาลัยโคโซกุ ประเทศญี่ปุ่น (พ.ศ. 2538-2541) ปัจจุบันเป็นมีตำแหน่งเป็นหัวหน้าฝ่ายพัฒนาและวิจัย ของโครงการของสำนักบริการเทคโนโลยีสารสนเทศภาครัฐ มีความสนใจเป็นพิเศษด้าน Network