

การดำเนินงานจดหมายเหตุมหาวิทยาลัยและ
ฐานข้อมูลจดหมายเหตุมหาวิทยาลัย

บทนี้มีวัตถุประสงค์ เพื่อนำเสนอการดำเนินการโครงการจดหมายเหตุมหาวิทยาลัย
ศิลปากร สรุปลักษณะและโครงสร้างของโปรแกรมสำเร็จรูป Mini-Micro CDS/ ISIS
Version 2.3 และ US MARC:AMC พร้อมทั้งนำเสนอทฤษฎี และแนวความคิดเกี่ยวกับการ
จัดเก็บและการค้นคืนสารสนเทศ โดยมีรายละเอียดของเนื้อหาที่จะกล่าวถึงดังนี้

1. ศึกษาการดำเนินงานโครงการจดหมายเหตุมหาวิทยาลัยศิลปากร
2. สรุปความเป็นมา และลักษณะการทำงานของ โปรแกรมสำเร็จรูป Mini-Micro
CDS/ISIS Version 2.3
3. สรุปประวัติ พัฒนาการ และลักษณะ พร้อมทั้งรายละเอียดของ US MARC:AMC
การศึกษาลักษณะการทำงานของ โปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS
Version 2.3 และ US MARC:AMC จะช่วยให้สามารถวางโครงสร้างระบบฐานข้อมูล และ
รูปแบบการแสดงผลข้อมูลได้ตรงตามจุดประสงค์
4. ศึกษาวรรณกรรมทางด้านการจัดเก็บและการค้นคืน และสรุปทฤษฎี หรือแนว
ความคิดที่สำคัญในเรื่องของการจัดเก็บและการค้นคืนข้อมูล โดยเน้นในเรื่องของการประเมินผล
การจัดเก็บและการค้นคืนสารสนเทศ ตัวแปรที่ใช้ในการประเมินผลการจัดเก็บและค้นคืนสารสนเทศ
คือ ค่า Recall และ ค่า Precision ซึ่งจะเป็นการทำความเข้าใจถึงวิธีการที่จะนำมาใช้
ในการประเมินผลระบบการจัดเก็บและการค้นคืนวัสดุจดหมายเหตุมหาวิทยาลัยศิลปากร จากฐาน
ข้อมูลที่สร้างขึ้น

การดำเนินการ โครงการจัดหาวัสดุศิลปากร

โครงการจัดหาวัสดุศิลปากร เป็นโครงการที่เริ่มจัดทำขึ้นเมื่อ พ.ศ. 2532 จากการที่ผู้บริหารของมหาวิทยาลัยได้ตระหนักถึงความจำเป็นในการรวบรวมบันทึกการดำเนินงาน และเหตุการณ์สำคัญๆ ของมหาวิทยาลัยศิลปากรไว้ เพื่อประโยชน์ต่อการอ้างอิงในการศึกษา ค้นคว้า วิจัย และการวางแผนพัฒนามหาวิทยาลัย โดยมีวัตถุประสงค์ในการดำเนินงาน ดังนี้ (โครงการจัดหาวัสดุศิลปากร, ม.ป.ป.)

1. เพื่อรวบรวม จัดเก็บรักษา จัดทำคู่มือช่วยค้นเอกสาร โสตทัศนวัสดุ และวัสดุที่มีความสำคัญต่องานจัดหาวัสดุศิลปากร
2. เพื่อรวบรวมและบันทึกเหตุการณ์ที่มีความสำคัญต่อประวัติ พัฒนาการและเกียรติประวัติของหน่วยงานและบุคลากร ในด้านต่างๆ ของมหาวิทยาลัย
3. เพื่อบริการจัดหาวัสดุศิลปากรแก่หน่วยงาน หรือบุคลากรภายในมหาวิทยาลัย เพื่อประโยชน์ในการบริหารงานและการศึกษาวิจัย

การจัดหาวัสดุจัดหาวัสดุของ โครงการจัดหาวัสดุศิลปากร

มี 3 วิธีคือ

1. การจัดซื้อ เป็นการจัดซื้อวัสดุจัดหาวัสดุของมหาวิทยาลัยศิลปากร ที่หน่วยงานราชการ หรือหน่วยงานเอกชนอื่นเป็นผู้จัดทำ แต่มีเนื้อหาที่เหมาะสม และสมควรเก็บเข้าโครงการจัดหาวัสดุศิลปากร ซึ่งฝ่ายพัฒนาทรัพยากรของห้องสมุดจะเป็นผู้ดำเนินการจัดซื้อให้ตามความต้องการของ โครงการฯ
2. การบริจาค วัสดุจัดหาวัสดุของ โครงการจัดหาวัสดุศิลปากร ส่วนใหญ่ได้รับบริจาคจากผู้ผลิต หรือผู้ที่เกี่ยวข้องกับมหาวิทยาลัยศิลปากรทั้งในอดีตและปัจจุบัน หรือจากอาจารย์ ข้าราชการ นักศึกษา เช่น รูปภาพ เอกสารประกอบการสอน เอกสารในการประชุมสัมมนาที่จัดขึ้นภายในมหาวิทยาลัยศิลปากร หนังสือที่ระลึก หนังสือรุ่น เป็นต้น
3. การได้รับในขั้นตอนการปฏิบัติราชการ เป็นวัสดุจัดหาวัสดุของมหาวิทยาลัยที่โครงการจัดหาวัสดุศิลปากรได้รับเมื่อ เอกสารนั้นเสร็จสิ้นการดำเนินงาน หรือยังอยู่ใน

ขั้นตอนการปฏิบัติงานภายในมหาวิทยาลัยศิลปากร และจัดส่งให้โครงการจดหมายเหตุมหาวิทยาลัย เป็นผู้จัดเก็บ ได้แก่ คำสั่งมหาวิทยาลัย รายงานการประชุมคณะดี สำนักมหาวิทยาลัย ประกาศ มหาวิทยาลัย

ประเภทของวัสดุจดหมายเหตุมหาวิทยาลัยศิลปากร

วัสดุจดหมายเหตุมหาวิทยาลัยศิลปากรที่ดำเนินการจัดเก็บ และให้บริการแบ่งออกเป็น 3 ประเภท ดังนี้

1. เอกสาร ประกอบด้วย

1.1 เอกสารที่หน่วยงานต่างๆ ของมหาวิทยาลัยจัดทำขึ้น เพื่อเผยแพร่ หรือใช้ประโยชน์ในหน่วยงาน เป็นเอกสารเกี่ยวกับนโยบาย ผลงานทางวิชาการ เช่น เอกสารประกอบการสอน บทความ หนังสือ รายงานการวิจัย เป็นต้น รวมทั้งรายงานการประชุม และเอกสารประกอบการประชุมของคณะกรรมการ อนุกรรมการต่างๆ ของมหาวิทยาลัยศิลปากร

1.2 ข่าวหรือบทความเกี่ยวกับมหาวิทยาลัย หรือบุคลากรของมหาวิทยาลัย ที่จัดทำโดยหน่วยงาน หรือองค์กรภายนอกมหาวิทยาลัย

1.3 เอกสารที่บุคลากรของมหาวิทยาลัยศิลปากรจัดทำขึ้นภายนอกมหาวิทยาลัย

2. โสตทัศนวัสดุ หมายถึง วัสดุจดหมายเหตุมหาวิทยาลัยศิลปากร ที่อยู่ในรูปแบบของวัสดุไมตินิมฟ์ แถบบันทึกเสียง วิทยุทัศน์ แผ่นผัง

3. วัตถุ หมายถึง วัตถุสิ่งของต่างๆ ที่เป็นผลงาน หรือลิขสิทธิ์ของมหาวิทยาลัยศิลปากร หรือบุคลากรในมหาวิทยาลัย

การดำเนินงานโครงการจดหมายเหตุมหาวิทยาลัยศิลปากร

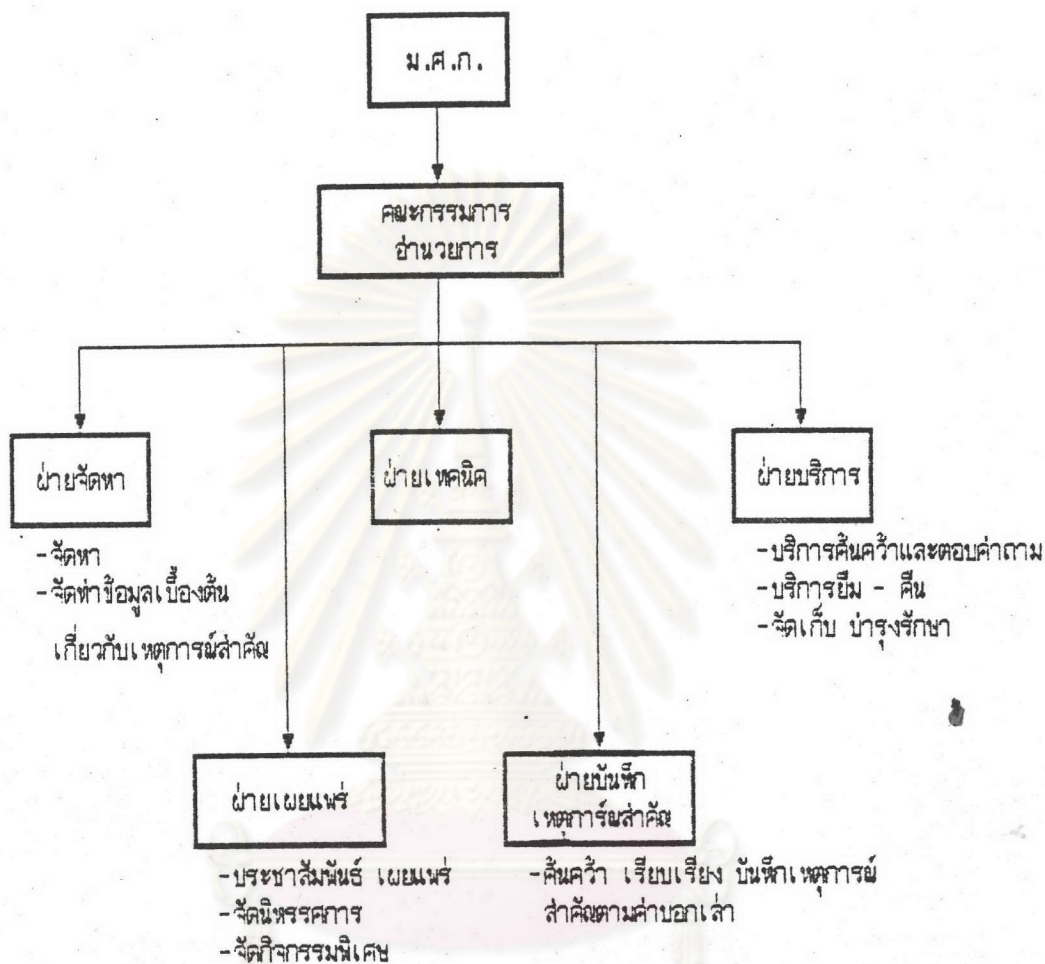
โครงการจดหมายเหตุมหาวิทยาลัยศิลปากร มีการแบ่งสายงานการบริหาร และกิจกรรมการดำเนินงานออกเป็น 5 ฝ่าย คือ

- สำคัญ
1. ฝ่ายจัดหา ทำหน้าที่จัดหา และจัดทำข้อมูลเบื้องต้นเกี่ยวกับเหตุการณ์
 2. ฝ่ายเทคนิค ทำหน้าที่จัดระบบ หมวดหมู่ และจัดทำคู่มือค้นคืน
 3. ฝ่ายบริการ ทำหน้าที่บริการช่วยการค้นคว้า และตอบคำถาม บริการให้ยืม รับผิดชอบ และจัดเก็บบำรุงรักษา
 4. ฝ่ายเผยแพร่ ทำหน้าที่ประชาสัมพันธ์ เผยแพร่ จัดนิทรรศการ จัดกิจกรรมพิเศษ
 5. ฝ่ายบันทึกเหตุการณ์สำคัญ ทำหน้าที่สอบค้น เรียบเรียง บันทึกเหตุการณ์สำคัญ และประวัติตามคำบอกเล่า

ลักษณะการดำเนินงาน โครงการจดหมายเหตุมหาวิทยาลัยศิลปากร สามารถเขียนเป็นแผนภูมิการบริหารงานได้ดังภาพประกอบที่ 1



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาพประกอบที่ 1: แผนภูมิการบริหารงานโครงการจดหมายเหตุมหาวิทยาลัยศิลปากร

การดำเนินงานตามโครงการนี้ ต้องอาศัยความร่วมมือจากหน่วยงานต่างๆ ของมหาวิทยาลัยศิลปากร โดยแต่ละหน่วยงานจะมีผู้แทนเข้าร่วมเป็นกรรมการ ในขณะทำงาน ฝ่ายต่างๆ โดยเฉพาะในส่วนของคณะกรรมการอำนวยการ จะมีรองอธิการบดีฝ่ายวิชาการ เป็นประธานกรรมการ ผู้ช่วยอธิการบดีวิทยาเขตพระราชวังสนามจันทร์ เป็นรองประธานกรรมการ นอกจากนี้ ผู้แทนจากคณะวิชาต่างๆ ของมหาวิทยาลัย คือ เลขานุการคณะฯ และประธานกรรมการฝ่ายต่างๆ ของโครงการจดหมายเหตุมหาวิทยาลัยศิลปากรเป็นกรรมการ คณะกรรมการอำนวยการนี้ มีหน้าที่กำหนดแนวทางการดำเนินงาน ให้เป็นไปตามนโยบายของมหาวิทยาลัย ให้คำปรึกษา เสนอแนะ จัดทำงบประมาณประจำปี และติดตามผลการปฏิบัติงานของโครงการ

ในปัจจุบันการจัดเก็บและค้นคืนวัสดุจดหมายเหตุมหาวิทยาลัยศิลปากร ยังไม่มีการจัดเก็บและค้นคืนด้วยเครื่องคอมพิวเตอร์ แต่มีการบันทึกข้อมูลของสิ่งพิมพ์อื่นๆ ของห้องสมุดลงเครื่องคอมพิวเตอร์ โดยใช้โปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS ซึ่งเป็นโปรแกรมที่ห้องสมุดส่วนใหญ่ในประเทศไทย ใช้ในการจัดเก็บและค้นคืนข้อมูล การสร้างฐานข้อมูลเพื่อจัดเก็บและค้นคืนวัสดุจดหมายเหตุมหาวิทยาลัยศิลปากร จึงใช้โปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS Version 2.3 ในการสร้างฐานข้อมูลด้วยความสะดวกในการแลกเปลี่ยนข้อมูลหรือใช้ฐานข้อมูลร่วมกัน

โปรแกรม CDS/ISIS

โปรแกรมสำเร็จรูป CDS/ISIS (เป็นคำย่อจาก Computerized Documentation System / Integrated Set of Information System) เป็นโปรแกรมที่สร้างขึ้นโดยองค์การแรงงานระหว่างประเทศ (International Labour Organization หรือ ILO) มีจุดประสงค์ที่จะแจกจ่ายให้กับประเทศต่างๆ ทั่วโลก จึงสร้างโปรแกรม CDS/ISIS ให้มีความยืดหยุ่นสูง ในปี พ.ศ. 2514 องค์การศึกษาวិทยาศาสตร์และวัฒนธรรมแห่งสหประชาชาติ (UNESCO) ได้นำโปรแกรม CDS/ISIS มาพัฒนาให้ส่วนเวรณปธิวิจิตรลงพิมพ์และพิมพ์ลงฉบับจริง และไมโครคอมพิวเตอร์ขนาด 16 บิตขึ้นไปได้ โดยเครื่องไมโครคอมพิวเตอร์ต้องมีหน่วยความจำหลักไม่น้อยกว่า 250 กิโลไบต์ และมีจานแม่เหล็กขนาดความจำไม่ต่ำกว่า 10 ล้านตัวอักษร โปรแกรม CDS/ISIS ที่ได้รับการพัฒนานี้ พัฒนาขึ้นด้วยภาษาปาสคาล (Pascal) ปฏิบัติงาน

ภายใต้ระบบปฏิบัติการ MS - DOS และเรียกโปรแกรมที่พัฒนาขึ้นนี้ว่า CDS/ISIS Mini-Micro Version และในปี พ.ศ. 2532 องค์การศึกษาวิทยาศาสตร์และวัฒนธรรมแห่งสหประชาชาติได้พัฒนาโปรแกรม CDS/ISIS นี้อีกครั้งหนึ่ง โดยเรียกว่า Mini-Micro CDS/ISIS Version 2.3 โดยมีการเปลี่ยนแปลง ให้โปรแกรมสามารถประมวลผลและจัดระบบฐานข้อมูลให้มีประสิทธิภาพสูงขึ้น (UNESCO, 1989)

หน้าที่ของโปรแกรมสำเร็จรูป CDS/ISIS

โปรแกรมสำเร็จรูป Mini - Micro CDS/ISIS มีหน้าที่หลักในการทำงานดังนี้

1. กำหนดและปรับปรุง โครงสร้างของฐานข้อมูล ได้ตามความต้องการ
2. บ้อนข้อมูลเข้าสู่ฐานข้อมูลที่กำหนดไว้แล้ว
3. ปรับปรุง แก้ไข และลบข้อมูลจากฐานข้อมูล
4. สร้างและบำรุงรักษา fast access file ที่ใช้ในการค้นคืนได้โดยอัตโนมัติ
5. ค้นคืนข้อมูลที่ต้องการจากฐานข้อมูล
6. แสดงผลของข้อมูลบนจอภาพ และทางเครื่องพิมพ์ได้ในรูปแบบที่ต้องการ
7. เรียงลำดับข้อมูล
8. ดูแลการรับและส่งข้อมูลจากฐานข้อมูลต่างๆ
9. พัฒนาการทำงานตามความต้องการเฉพาะของผู้ใช้

โปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS มีระบบปฏิบัติงานหลักของโปรแกรมแบ่งออกเป็น 2 ส่วน คือ (สุวรรณ ทงสิสุขใส, 2533: 2)

1. USER SERVICES ใช้ในการบันทึก แก้ไขปรับปรุง แสดงรายงาน ค้นคืนข้อมูล สั่งพิมพ์ข้อมูล จัดเก็บคำค้นคืน ประกอบด้วยโปรแกรมย่อย ดังนี้
 - 1.1 ISISENT ใช้ในการทำงานเกี่ยวกับการบำรุงรักษาฐานข้อมูล เช่น การบันทึกข้อมูลระเบียนใหม่ การแก้ไขข้อมูล
 - 1.2 ISISRET ใช้ในการค้นคืนข้อมูล และแสดงข้อมูล
 - 1.3 ISISPRT ใช้ในการสั่งพิมพ์รายงาน
 - 1.4 ISISINV ใช้ในการดำเนินงานเกี่ยวกับแฟ้มข้อมูลดัชนี

2. SYSTEM SERVICES ใช้ในการจัดการฐานข้อมูล เพื่อให้สามารถสร้างฐานข้อมูล การทำงานที่เกี่ยวข้องกับระบบ และการแลกเปลี่ยนข้อมูลกับฐานข้อมูลอื่น ประกอบด้วยโปรแกรมย่อย ดังนี้

2.1 ISISDEF ใช้ในการสร้างฐานข้อมูล เช่น เพิ่มข้อมูลหลัก เพิ่มข้อมูลดัชนี แผ่นป้อนข้อมูล การสร้างรูปแบบการแสดงรายงาน

2.2 ISISUTL ใช้ในการสร้างและแก้ไขเมนู และแผ่นป้อนข้อมูลของระบบ พิมพ์เมนู กำหนดคุณลักษณะของจอภาพ พิมพ์ข้อความบอกบท หรือคำสั่งโต้ตอบ

2.3 ISISXCH ใช้ในการถ่ายข้อมูลจากฐานข้อมูลอื่น เข้าสู่ฐานข้อมูลที่ใช้งาน และถ่ายข้อมูลออกจากฐานข้อมูลที่ใช้งานไปใช้ในฐานข้อมูลอื่น

2.4 ISISPAS ใช้ในการดำเนินงานโปรแกรมที่เขียนเพิ่มเติม เพื่อใช้ในการประมวลผลเฉพาะอย่าง

ลักษณะของโปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS

โปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS นี้ นอกจากจะอำนวยความสะดวกต่อผู้ใช้ ในการติดตั้ง โปรแกรมไว้ใช้งานกับเครื่อง ไมโครคอมพิวเตอร์ขนาด 16 บิตขึ้นไป ซึ่งมีใช้กันอย่างแพร่หลายแล้วนั้น โปรแกรมที่พัฒนาขึ้นนี้ ยังคงมีฟังก์ชันคล้ายกับโปรแกรมที่ใช้บนเครื่องคอมพิวเตอร์ขนาดใหญ่ด้วย และพัฒนาให้มีความคล่องตัวในการใช้งานแบบเครื่อง ไมโครคอมพิวเตอร์โดยมีลักษณะเป็น Menu-driven กล่าวคือ ผู้ใช้สามารถเลือกประมวลผลข้อมูลได้อย่างสะดวก โดยผ่านทางเมนู(Menu) ของระบบ ทำให้ใช้งานได้สะดวก และผู้ใช้ไม่จำเป็นต้องมีพื้นฐานทางคอมพิวเตอร์มากนัก ลักษณะที่สำคัญของ โปรแกรม คือ

1. การเชื่อมเมนูย่อยโดยมีลำดับชั้น ลักษณะการเรียกโปรแกรมย่อยผ่านเมนู เป็นสิ่งที่ทำให้โปรแกรม Mini-Micro CDS/ISIS มีลักษณะง่ายต่อการใช้งาน การรวมโปรแกรมย่อยทั้งหมดเข้าในเมนูหลักอันเดียวกันอย่างมีระบบ จะทำให้ผู้ปฏิบัติงานสามารถใช้งานได้ง่ายขึ้น โดยโปรแกรมจะเชื่อมเมนูย่อยต่างๆ เข้าด้วยกันอย่างมีลำดับ ผู้ใช้สามารถประมวลผลโปรแกรมเหล่านี้ได้โดยเพียงเลือกผ่านรายการระบบหลัก (Main Menu) ที่ชื่อว่า xXISI

2. สามารถประมวลผลได้ทั้ง ในลักษณะแบทช์และออนไลน์ ให้ผู้ใช้สามารถเลือกใช้ตามความเหมาะสมกับลักษณะงานและอุปกรณ์ที่มีอยู่

3. ระบบจัดการฐานข้อมูลได้ออกแบบเป็นพิเศษ สำหรับการจัดเก็บข้อมูลประเภทตัวอักษรที่ไม่จำกัดขนาด

4. สามารถใช้กับการจัดการฐานข้อมูล ที่มีลักษณะโครงสร้างของข้อมูลที่แตกต่างกันได้

5. การทำดัชนี สามารถทำได้ทั้งในลักษณะดัชนีควบคุม ซึ่งช่วยให้การเรียกใช้ข้อมูลจากตารางคำศัพท์ สามารถทำได้อย่างมีประสิทธิภาพ หรือทำดัชนีในลักษณะไม่ควบคุมคำศัพท์ หรือที่เรียกว่า Natural Language ซึ่งนำไปจัดทำ Inverted file เพื่อการค้นคืนข้อมูล

6. การค้นคืนข้อมูลจากฐานข้อมูลสามารถทำได้ทั้งแบบออนไลน์ และแบบแบทช์ โดยการค้นคืนแบบออนไลน์ อำนวยความสะดวกกับผู้ใช้ในการติดต่อกับฐานข้อมูลโดยตรง สามารถค้นคืนข้อมูลในลักษณะเจาะจง หรือกลุ่มของคำที่มีลักษณะใกล้เคียงกัน ตลอดจนสามารถค้นคืนโดยการผสมคำด้วยเทคนิค Boolean นอกจากนี้ยังสามารถทำระบบแจกจ่ายเอกสารตามที่กำหนด หรือ SDI ได้โดยวิธีการประมวลผลแบบแบทช์

7. การกำหนดคำสั่งรูปแบบการพิมพ์ที่ออกแบบเป็นพิเศษ ช่วยให้มืออิสระในการแสดงรูปแบบของผลลัพธ์ ไม่ว่าจะแสดงบนจอภาพ หรือพิมพ์ออกทางเครื่องพิมพ์

นอกจากนี้ยังมีการกำหนด Menu Jump Keys เพื่อใช้กระโดดไปยังรายการระบบที่ต้องการได้อย่างรวดเร็ว คือ

- F2 - เลือกรายการหลัก (xXISI)
- F3 - เลือกรายการระบบการเรียกแสดงผลข้อมูล (xXGEN)
- F4 - เลือกรายการระบบของการบันทึกข้อมูล (xXE1)
- F5 - เลือกรายการระบบของการเรียงลำดับและการพิมพ์ (xXPRT)

โปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS ยังมีลักษณะพิเศษ อื่นๆ อีก คือ (สุวรรณ ทงสิทธิ์, 2533: 1)

1. จัดเก็บข้อมูลไว้เป็นส่วนๆ เรียกว่า เขตข้อมูล ใช้ตัวเลขเป็นรหัสแทนเขตข้อมูล แต่ละเขตข้อมูล แบ่งเป็นเขตข้อมูลย่อยได้อย่างไม่จำกัด และมีข้อมูลได้มากกว่า 1 ครั้ง (Repeatable field)

2. เก็บข้อมูลแบบความยาวไม่แน่นอน คือ เก็บข้อมูลตามที่เป็นจริง ไม่เว้นที่ว่างสำหรับเขตข้อมูลที่ไม่มีข้อมูล ทำให้ประหยัดเนื้อที่ในหน่วยความจำ
3. สามารถกำหนดคำที่ไม่ต้องการให้เรียงได้ เช่น ไม่ให้เรียง Article ที่เป็นคำแรกของชื่อเรื่อง โดยใช้เครื่องหมาย < > คร่อมคำที่ไม่ต้องการให้เรียง
4. สั่งให้เรียงตัวเลขตามคำอ่านได้ โดยแสดงวิธีอ่านไว้ในเครื่องหมาย < > เช่น เลข 12 จะอ่านเป็น twelve และเรียงไว้ที่ t จะสั่งโดยใช้วิธี <12=twelve>
5. เรียงคำย่อไว้ก่อนคำที่มีตัวสะกด เช่น WHO จะเรียงไว้ตอนต้นของคำที่ขึ้นต้นด้วย W ก่อนคำที่ขึ้นต้นด้วย W ที่ตามด้วยตัวสะกด วิธีสั่งจะใช้ <WHO=W H O>
6. กำหนดให้ทำดัชนีเพื่อใช้ค้นคืนได้ตามต้องการ เช่น เก็บทั้งเขตข้อมูล หรือทุกเขตข้อมูลย่อย หรือเฉพาะบางเขตข้อมูลย่อย หรือทุกคำ หรือเฉพาะคำใดคำหนึ่ง นำมาเก็บเป็นคำหรือวลีที่ใช้ในการค้นคืน

ข้อจำกัดของ โปรแกรมสำเร็จรูป Mini-Micro CDS/ISIS

1. จำนวนของฐานข้อมูล	ไม่จำกัด
2. จำนวนระเบียบที่มีได้สูงสุดในฐานข้อมูล	16 ล้านตัวอักษร
3. ขนาดของระเบียบที่ใหญ่ที่สุด	8,000 อักขระ
4. จำนวนเขตข้อมูลที่สามารถกำหนดใน FDT ได้สูงสุด	200 เขตข้อมูล (ไม่รวมเขตข้อมูลซ้ำ)
5. จำนวนบรรทัดใน FST ที่มีได้สูงสุด	200 บรรทัด
6. ขนาดของเขตข้อมูลสูงสุด	8,000 อักขระ
7. จำนวนเขตข้อมูลที่บรรจุใน 1 หน้าของแผ่นงาน	19 บรรทัด
8. จำนวนหน้าสูงสุดของแผ่นงาน	20 หน้า
9. ขนาดของรูปแบบการพิมพ์ (display format)	4,000 อักขระ
10. ขนาดของ Stopwords file	799 คำ

การดำเนินการสร้างฐานข้อมูลในเครื่องคอมพิวเตอร์ เพื่อจัดเก็บข้อมูลรายละเอียดของสิ่งพิมพ์ หรือวัสดุต่างๆ นั้น จะต้องมีการกำหนดโครงสร้างระเบียบของฐานข้อมูลที่ใช้ในการบันทึกข้อมูล โครงสร้างระเบียบที่ใช้ควรจะเป็นโครงสร้างที่ได้มาตรฐาน ซึ่งจะช่วยให้สามารถ

แลกเปลี่ยนถ่ายเทข้อมูลกับหน่วยงาน หรือสถาบันอื่นๆ ได้ง่าย

การลงรายการของวัสดุจดหมายเหตุมหาวิทยาลัยศิลปากรที่ดำเนินการสร้างฐานข้อมูลนี้ จะใช้มาตรฐานการลงรายการวัสดุจดหมายเหตุที่เครื่องคอมพิวเตอร์สามารถอ่านได้ ของสมาคม นักจดหมายเหตุอเมริกัน (US MARC: AMC) ในการบันทึกข้อมูลรายละเอียดของวัสดุจดหมายเหตุ

มาตรฐานการลงรายการวัสดุจดหมายเหตุที่เครื่องคอมพิวเตอร์สามารถอ่านได้ ของสมาคมนัก
จดหมายเหตุอเมริกัน (US MARC: Archival and Manuscripts Control)

มาตรฐานการลงรายการวัสดุจดหมายเหตุที่เครื่องคอมพิวเตอร์สามารถอ่านได้ ของ สมาคมนักจดหมายเหตุอเมริกัน (US MARC: AMC) คือ มาตรฐานการลงรายการของวัสดุ จดหมายเหตุ ที่อ้างอิงรูปแบบโครงสร้างมาตรฐาน US MARC ซึ่งเป็นรูปแบบมาตรฐานสำหรับการลงรายการทางบรรณานุกรมของสิ่งตีพิมพ์ และ โสัดทัศน์วัสดุต่างๆ ที่เครื่องคอมพิวเตอร์ สามารถอ่านได้ และใช้เป็นมาตรฐานในการแลกเปลี่ยนข้อมูลระหว่างฐานข้อมูลคอมพิวเตอร์ (Sahli, 1986)

ประวัติและพัฒนาการของ US MARC:AMC Format.

Shali (1985, 1986) ซึ่งเป็นคณะกรรมการของ National Historical Publication and Records Commission หรือ NHPRC และเป็นผู้เชี่ยวชาญด้านจดหมายเหตุ โดยเฉพาะเกี่ยวกับการนำเทคโนโลยีมาใช้กับจดหมายเหตุ ได้สรุปถึงประวัติ พัฒนาการ ของ US MARC:AMC ว่า ชื่อเต็มของ US MARC: AMC Format คือ US MARC: Archival and Manuscripts Control Format มีพื้นฐานโครงสร้างร่วมกับมาตรฐาน US MARC ซึ่งพัฒนา ขึ้นโดย หอสมุดรัฐสภาอเมริกัน และนำออกเผยแพร่ในราวปี ค.ศ. 1973 แต่ไม่ได้รับการยอมรับ จากวงการนักจดหมายเหตุ เพราะขอบเขตของ โครงสร้างระเบียบมีขีดจำกัดจนไม่สามารถแสดง ให้เห็นถึงลักษณะและรายละเอียดของข้อมูล และมีโครงสร้างระเบียบไม่เพียงพอต่อความต้องการ ของผู้ใช้ ดังนั้นสมาคมนักจดหมายเหตุอเมริกัน (Society of American Archivists) จึงได้ ก่อตั้งคณะกรรมการที่เรียกว่า National Information System Task Force (NISTF) ขึ้น เพื่อศึกษาปัญหาที่เกิดขึ้น NISTF ได้พิจารณาพบว่า ระบบใดระบบหนึ่งเพียงระบบเดียวไม่สามารถ

สนองตอบความต้องการของผู้ใช้จดหมายเหตุได้อย่างครบถ้วน และรูปแบบโครงสร้างระเบียบควรมีรูปแบบที่ง่าย สะดวกต่อการแลกเปลี่ยนข้อมูลจดหมายเหตุ สามารถใช้งานได้กับเครื่องคอมพิวเตอร์ที่มีลักษณะแตกต่างกันทั้ง ฮาร์ดแวร์ หรือ ซอฟต์แวร์ สามารถประยุกต์ใช้ในการปฏิบัติงานที่ทำด้วยมือได้ และควรสามารถแลกเปลี่ยนกับ โครงสร้างที่ได้มาตรฐานอื่นได้ด้วย

หลังจากการสำรวจความต้องการของผู้ใช้เพื่อพัฒนารูปแบบโครงสร้างระเบียบใหม่แล้ว NISTF ได้ตัดสินใจว่าการนำรูปแบบเดิมที่มีอยู่แล้วมาปรับปรุงจะเป็นวิธีการที่ประหยัดที่สุด ดังนั้น NISTF จึงได้นำเอารูปแบบโครงสร้าง US MARC Manuscripts มาปรับปรุง โดยร่วมมือกับ Research Libraries Group (RLG) นำรูปแบบโครงสร้างระเบียบที่ได้รับการปรับปรุงแล้ว เข้าไปเป็นส่วนประกอบในระบบออนไลน์ ของ Research Libraries Information Network (RLIN) ทั้ง NISTF และ RLG ได้ทำงานควบคู่ไปกับหอสมุดรัฐสภาอเมริกัน เพื่อพัฒนารูปแบบโครงสร้างระเบียบที่เหมาะสมจนในต้นปี ค.ศ. 1983 จึงได้กำหนดรูปแบบโครงสร้างระเบียบขั้นใหม่ เรียกว่า Archival and Manuscripts Control Format และได้รับการรับรองจากสมาคมนักจดหมายเหตุอเมริกัน หอสมุดรัฐสภาอเมริกัน และคณะกรรมการซึ่งเป็นตัวแทน ในการกำหนดรูปแบบทางบรรณานุกรมที่สามารถอ่านได้ด้วยเครื่องคอมพิวเตอร์ ของสมาคมห้องสมุดอเมริกัน ต่อมาในช่วงระหว่างปี ค.ศ. 1983-1984 ได้มีการพัฒนารูปแบบโครงสร้างอีกครั้งหนึ่ง เป็นการปรับปรุงความหมายของเขตข้อมูล(field) และเขตข้อมูลย่อย(subfield) พร้อมทั้งให้ตัวอย่างในการลงรายการด้วย

นอกจากนั้น Sahli(1985) ยังอธิบายการใช้ US MARC:AMC ว่ารูปแบบโครงสร้างระเบียบของ AMC Format มีลักษณะเหมือนกับรูปแบบโครงสร้างของ US MARC โดยมีส่วนที่แตกต่างกันเพียงบางส่วน ซึ่งกำหนดไว้เพื่อการจัดเก็บข้อมูลวัสดุจดหมายเหตุโดยเฉพาะ ในแต่ละระเบียบของ Archival and Manuscripts Control Format จะประกอบไปด้วย

- Leader
- Directory
- Variable Control Field
- Variable Data Field

ในส่วนของ Leader และ Directory จะใช้ในการเก็บข้อมูลเกี่ยวกับระเบียบนั้นๆ และเป็นส่วนที่เครื่องคอมพิวเตอร์ใช้ในการค้นคืนข้อมูลในระเบียบ

ส่วน Variable Control Field จะให้ข้อมูลเกี่ยวกับหมายเลข ที่ใช้ควบคุมระเบียบ วัน เวลาที่มีการเปลี่ยนแปลงข้อมูลครั้งล่าสุด รูปร่างลักษณะของวัสดุ หรือเก็บรหัสต่างๆ ที่ใช้ประโยชน์ในการค้นคืนข้อมูล

Variable Data Field ประกอบด้วยเขตข้อมูลทั้งสิ้น 77 เขตข้อมูล เป็นส่วนที่ใช้ในการบันทึกข้อมูลรายละเอียดของวัสดุจดหมายเหตุ ส่วนประกอบของ Variable Data Field จะมีหมายเลข Tag ประจำแต่ละเขตข้อมูล เป็นหมายเลขระหว่าง 010 ถึง 886 ในแต่ละเขตข้อมูล จะประกอบด้วย อักขระ 2 ตัว ที่เรียกว่า Indicator ซึ่งจะเป็นตัวสรุปเนื้อหาของข้อมูลที่มีอยู่ต่อจาก Indicator ในแต่ละเขตข้อมูลจะมีเขตข้อมูลย่อย (Subfield) มีตั้งแต่ 1-20 เขตข้อมูลย่อย ซึ่งจะเก็บข้อมูลต่างๆ ของเอกสาร เช่น ชื่อเรื่อง ช่วงเวลา คำตัดขึ้น เป็นต้น อย่างน้อยที่สุดในแต่ละเขตข้อมูลจะต้องมี 1 เขตข้อมูลย่อย

รายละเอียดของเขตข้อมูลในการลงรายการของ US MARC:AMC ได้แสดงไว้ในตารางที่ 1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 : รายละเอียดเขตข้อมูลของ US MARC: AMC



เขตข้อมูล	ชื่อเขตข้อมูล
001	Control number
002	Subrecord map of directory
005	Date and time of latest transaction
007/00	Category of material
007/01	Specific material designation
007/02	Original versus reproduction aspect
007/03	Porality (microforms)
007/04	Dimensions (microforms)
007/05-08	Reduction ratio
007/09	Color (microforms)
007/10	Emulsion on film (microforms)
007/11	Generation
007/12	Base of film (microforms)
008/00-05	Date entered on file
008/06	Type of date code
008/07-10	Date 1
008/11-14	Date 2
008/15-17	Place of publication, production, or execution code
008/18-22	Undefined
008/23	Form of reproduction code

ตารางที่ 1 : รายละเอียดเขตข้อมูลของ US MARC: AMC (ต่อ)

เขตข้อมูล	ชื่อเขตข้อมูล
008/24-34	Undefined
008/35-37	Language code
008/38	Modified record code
008/39	Cataloging source code
010	Library of congress control number
035	Local system control number
039	Level of bibliographic control and coding detail
040	Cataloging source
041	Language code
043	Geographic classification code
045	Chronological code or date/time
052	Geographic classification code
066	Character sets present
072	Subject category code
09x	Local call numbers
100	Main entry - personal name
110	Main entry - corporate name
111	Main entry - uniform title heading
130	Main entry - conference or meeting
240	Uniform title
242	Translation of title by cataloging agency
243	Uniform title, collective

ตารางที่ 1: รายละเอียดเขตข้อมูลของ US MARC: AMC (ต่อ)

เขตข้อมูล	ชื่อเขตข้อมูล
245	Title statement
260	Publication, distribution, etc.(imprint)
300	Physical description
340	Medium
351	Organization and arrangement
500	General note
502	Dissertation note
505	Contents note (formatted)
506	Restrictions on access
510	Citation note (brief form)/references
520	Summary, abstract, annotation, scope, etc
521	Users/intend audience note
524	Preferred citation of described materials
530	Additional physical form available note
533	Reproduction note
535	Location of originals/duplicates
540	Terms governing use and reproduction
541	Immediate source of acquisition
544	Location of associated materials
545	Biographical or historical note
546	Language note
555	Cumulative index/finding aids note
561	Provenance

ตารางที่ 1: รายละเอียดเขตข้อมูลของ US MARC: AMC (ต่อ)

เขตข้อมูล	ชื่อเขตข้อมูล
562	Copy and version identification
565	Case file characteristics note
580	Linking entry complexity note
581	Publications note
583	Actions
584	Accumulation and frequency of use
59x	Local notes
600	Subject added entry - personal name
610	Subject added entry - corporate name
611	Subject added entry - conference or meeting
630	Subject added entry - uniform title heading
650	Subject added entry - topical heading
651	Subject added entry - geographic name
655	Genre/form heading
656	Index term - occupation
657	Index term - function
69x	Local subject added entries
700	Added entry - personal name
710	Added entry - corporate name
711	Added entry - conference or meeting
730	Added entry - uniform title heading

ตารางที่ 1: รายละเอียดเขตข้อมูลของ US MARC: AMC (ต่อ)

เขตข้อมูล	ชื่อเขตข้อมูล
740	Added entry - title traced differently
752	Added entry - place of publication or production
773	Host item entry
851	Location
870	Variant personal name
871	Variant corporate name
872	Variant conference or meeting name
873	Variant uniform title heading
880	Alternate graphic representation
886	Foreign MARC information field

การใช้คอมพิวเตอร์ในการบริหารงานจดหมายเหตุมหาวิทยาลัย

การนำคอมพิวเตอร์เข้ามาใช้ ในการดำเนินงานด้านจดหมายเหตุมหาวิทยาลัยของมหาวิทยาลัยในต่างประเทศ โปรแกรมที่มหาวิทยาลัยต่างๆ นำมาทดสอบและใช้นั้น มีทั้งโปรแกรมที่เขียนขึ้นด้วยภาษาคอมพิวเตอร์ และโปรแกรมที่เป็นโปรแกรมสำเร็จรูป บางสถาบันในระยะแรก พัฒนาระบบโดยการเขียนโปรแกรมด้วยภาษาคอมพิวเตอร์ และปรับเปลี่ยนมาใช้โปรแกรมสำเร็จรูปในภายหลัง ซึ่งผู้ปฏิบัติงานด้านจดหมายเหตุคาดว่า จะสามารถสนองตอบความต้องการของผู้ปฏิบัติงาน และผู้ใช้บริการได้ดี หอจดหมายเหตุมหาวิทยาลัยในต่างประเทศที่นำคอมพิวเตอร์เข้ามาช่วยในการดำเนินงานจดหมายเหตุมหาวิทยาลัย เช่น

มหาวิทยาลัยลิเวอร์พูล (Liverpool University)

หอจดหมายเหตุมหาวิทยาลัยลิเวอร์พูล เขียนโปรแกรมสร้างระบบฐานข้อมูลเพื่อใช้ในการควบคุมการดำเนินงานเกี่ยวกับเอกสาร ในช่วงระหว่างปี ค.ศ.1978 - ค.ศ.1984 ภาษาคอมพิวเตอร์ที่ใช้เป็นภาษาโคบอล ฐานข้อมูลที่สร้างขึ้นมีชื่อว่า CMF โดยใช้เครื่องคอมพิวเตอร์เมนเฟรมรุ่น ICL 1906S และใช้ระบบปฏิบัติงานที่ชื่อ George 4 ระบบ CMF ประกอบไปด้วยโปรแกรมย่อย 3 โปรแกรม คือ โปรแกรมที่ใช้ในการจัดเรียงข้อมูล ได้แก่ เรียงตามหมายเลขอ้างอิง (Reference code order) เรียงตามวันที่ที่ส่งมอบเอกสาร (Disposal date order) และเรียงตามคำดัชนีหัวเรื่อง (Subject-index keyword order) ต่อมาเครื่องคอมพิวเตอร์รุ่น ICL เลิกผลิต และมีเครื่อง IBM เข้ามาแทนทำให้หอจดหมายเหตุมหาวิทยาลัยลิเวอร์พูลเริ่มนำระบบโปรแกรมสำเร็จรูปมาใช้ โดยได้ใช้ระบบ RAPPORT ซึ่งเป็นระบบจัดการฐานข้อมูล (Data base management systems) กับเครื่อง IBM 3083 เป็นระบบฐานข้อมูลที่สามารถเชื่อมโยงแฟ้มข้อมูลต่างๆ ได้ดี และผู้สร้างฐานข้อมูลสามารถกำหนดโครงสร้างของแฟ้มข้อมูลได้หลายวิธี ส่วนการบริการข้อมูลจดหมายเหตุมหาวิทยาลัย ข้อมูลต่างๆ จะถูกส่ง ไปให้บริการตามคณะวิชาต่างๆ ในมหาวิทยาลัย การค้นคืนเป็นไปแบบปฏิสัมพันธ์ (Interactive) และผลการค้นคืนสามารถจัดเก็บแยกแฟ้มข้อมูลใหม่ หรือสั่งพิมพ์ออกจากเครื่องพิมพ์ก็ได้ (Cook, 1986: 73-74)

มหาวิทยาลัยกลาสโกว์ (Glasgow University)

ระบบแรกที่หอจดหมายเหตุมหาวิทยาลัยกลาสโกว์พัฒนาขึ้นในปี ค.ศ.1980 คือระบบ PARCH ซึ่งหลังจากนั้นจึงได้มีการนำเอาระบบสำเร็จรูปที่ใช้จัดการงานด้านบรรณานุกรมที่มีอยู่แล้วมาทดลองใช้ คือ ระบบ FAMULUS เนื่องจากเป็นระบบที่มีคุณสมบัติตามที่หอจดหมายเหตุส่วนใหญ่ต้องการ นอกจากนั้นระบบ FAMULUS ยังสามารถใช้ได้ง่าย ผู้ที่ไม่มีผู้เชี่ยวชาญเกี่ยวกับระบบคอมพิวเตอร์ หรือบุคคลทั่วไปสามารถใช้งานได้ ระบบ FAMULUS ประกอบด้วยโปรแกรมย่อย 8 โปรแกรม คือ

1. โปรแกรมการจัดเรียง (Sort)
2. โปรแกรมเชื่อมโยง (Cross-reference)

3. โปรแกรมการปรับปรุงแฟ้มข้อมูลให้ทันสมัย (Update)
4. โปรแกรมการแก้ไข (Edit)
5. โปรแกรมการค้นคืน (Search)
6. โปรแกรมการคำนวณ (Count)
7. โปรแกรมการตรวจสอบ (Validate)
8. โปรแกรม Multisearch

หอจดหมายเหตุมหาวิทยาลัยกลาสโกว์ ใช้บริการคอมพิวเตอร์จากคอมพิวเตอร์ ส่วนกลางของมหาวิทยาลัย จึงทำให้ฐานข้อมูลจดหมายเหตุมหาวิทยาลัยกลาสโกว์สามารถ ค้นคืนได้จากจุดต่างๆภายในมหาวิทยาลัยอย่างทั่วถึง (Cook, 1986: 104-105)

วิทยาลัยเซนต์จอห์น (St. John's College)

หอจดหมายเหตุวิทยาลัยเซนต์จอห์นใช้ระบบ GOSLING มาตั้งแต่ปี ค.ศ. 1977 ระบบ GOSLING เป็น VERSION ย่อยของ GOS สำหรับใช้งานบนเครื่องไมโครคอมพิวเตอร์ เป็นระบบที่พัฒนาขึ้นด้วยภาษาฟอร์แทรน โดย Museum Documentation Association (MDA) หอจดหมายเหตุวิทยาลัยเซนต์จอห์น นำระบบ GOSLING มาใช้เพื่อการจัดหมวดหมู่วัสดุจดหมายเหตุ มหาวิทยาลัย ที่ต้องมีการแยกหมวดหมู่อย่างละเอียด นอกจากนั้นแบบฟอร์มในการป้อนข้อมูล ยังได้รับการออกแบบเป็นพิเศษ เพื่อใช้บันทึกข้อมูลต่างๆ ของจดหมายเหตุ ส่วนการค้นคืน สามารถค้นคืนข้อมูลเขตข้อมูลต่างๆ หรือหลายเขตข้อมูลรวมกันในลักษณะออนไลน์ และสามารถ จัดทำคู่มือเป็นรูปเล่มได้ (Cook, 1986: 81)

มหาวิทยาลัยแห่งรัฐมิชิแกน (Michigan State University)

มหาวิทยาลัยแห่งรัฐมิชิแกนได้พัฒนาระบบ MicroMARC:amc ในปี ค.ศ. 1983 จนถึง ค.ศ. 1986 โดยได้รับการสนับสนุนด้านงบประมาณจาก National Historical Publications and Records Commission (NHPRC) การพัฒนาระบบ MicroMARC: amc มีจุดประสงค์ให้เป็นระบบอัตโนมัติสำหรับการปฏิบัติงาน และควบคุมวัสดุจดหมายเหตุ และต้นฉบับ ตัวเขียนได้ตลอดวงจรชีวิตของเอกสาร เป็นระบบที่สามารถทำงานกับระบบปฏิบัติงานบนเครื่อง

ไมโครคอมพิวเตอร์ได้ ทำให้หोजดหมายเหตุมหาวิทยาลัย นิพริภณณ์ที่ สถาบันจดหมายเหตุ ที่มีขนาดเล็กและมีอุปกรณ์คอมพิวเตอร์น้อยสามารถนำระบบนี้ไปใช้ได้ นอกจากนั้นระบบ MicroMARC: amc ได้รับการออกแบบให้ใช้ง่าย โดยมีลักษณะเป็นเมนูให้เลือก โครงสร้างของระเบียบที่ใช้ในระบบ MicroMARC: amc ใช้มาตรฐาน US MARC: AMC ทำให้สามารถถ่ายเทข้อมูลในระบบ MicroMARC: amc นี้ไปยังฐานข้อมูลอื่นได้ ในปี ค.ศ.1988 คณะผู้ทำงานของโครงการระบบ MicroMARC: amc ได้พัฒนาให้ระบบสามารถแปลงข้อมูลที่มีอยู่ในระบบให้ไปอยู่ในรูปแบบรหัส ASCII ได้ จึงทำให้ข้อมูลต่างๆ ที่มีอยู่สามารถเชื่อมโยงกับฐานข้อมูลบน ไมโครคอมพิวเตอร์ เครื่องอื่นๆ ที่สามารถรับรหัส ASCII

ระบบ MicroMARC: amc เป็นระบบจัดการฐานข้อมูลที่มีประสิทธิภาพสามารถเข้าถึงข้อมูลในฐานข้อมูลได้หลายวิธี โปรแกรมการค้นคืนของระบบทำงานได้อย่างรวดเร็ว และสามารถส่งต่อข้อมูลไปยังโปรแกรมการออกรายงาน

นอกจากหोजดหมายเหตุมหาวิทยาลัยแห่งรัฐมิชิแกน และสถาบันที่ร่วมทำการทดสอบรวม 5 สถาบัน ได้แก่ มหาวิทยาลัย Georgetown สมาคมประวัติศาสตร์แห่งรัฐอินเดียนา ห้องสมุดทางการเกษตรแห่งชาติ มหาวิทยาลัย Louisville และมหาวิทยาลัย Case Western Reserve ซึ่งได้พัฒนาและใช้ระบบนี้แล้ว ยังมีสถาบันต่างๆ นำไปใช้มากกว่า 60 สถาบัน รวมทั้งหोजดหมายเหตุมหาวิทยาลัยและวิทยาลัย หोजดหมายเหตุของรัฐ สถาบันทางประวัติศาสตร์ มุณินธิต่างๆ และหोजดหมายเหตุทางธุรกิจ จึงถือได้ว่าเป็นระบบที่มีสถาบันต่างๆ นิยมใช้มากระบบหนึ่ง (Honhart, 1989)

มหาวิทยาลัยอิลลินอยส์ (University of Illinois)

มหาวิทยาลัยอิลลินอยส์ใช้ระบบ PARADIGM ในการดำเนินงานด้านจดหมายเหตุ มหาวิทยาลัย PARADIGM เป็นระบบประมวลผลข้อมูลอัตโนมัติที่พัฒนาขึ้นในปี ค.ศ.1971 โดยมหาวิทยาลัยอิลลินอยส์ เพื่อใช้ในการดำเนินงาน และควบคุมหัวเรื่องของจดหมายเหตุมหาวิทยาลัย William J. Maher นักจดหมายเหตุมหาวิทยาลัยแห่งมหาวิทยาลัยอิลลินอยส์ สรุปรวตฤประสงค์ของระบบ PARADIGM ว่าเพื่อให้คอมพิวเตอร์ช่วยในการจัดเรียงจดหมายเหตุมหาวิทยาลัยที่มีความละเอียด และช่วยควบคุมคำสำคัญที่คัดเลือกจากจดหมายเหตุมหาวิทยาลัยที่มีจำนวนมาก ซึ่ง

คำสำคัญของจดหมายเหตุมหาวิทยาลัยร้อยละ 76 เป็นชื่อเฉพาะ (Proper name) โดยคัดเลือกมาจากบัตรควบคุม เครื่องมือช่วยค้น และจากเนื้อหาของเอกสารระบบ PARADIGM จะช่วยให้สามารถเข้าถึงเอกสารได้ละเอียด แต่ PARADIGM เป็นระบบที่ไม่มีศัพท์สัมพันธ์ สามารถทำงานในระบบออนไลน์ และใช้ร่วมกับระบบบูรณาการ (integrated system) ได้ (Cook, 1986; Berner, 1983)

โปรแกรมคอมพิวเตอร์อื่นๆ ที่หอจดหมายเหตุมหาวิทยาลัยนำไปใช้ในการดำเนินงาน ได้แก่

STATUS เป็นโปรแกรมสำเร็จรูปสำหรับการค้นคืนสารนิเทศ ซึ่งเมื่อเปรียบเทียบกับ FAMULUS แล้ว STATUS จะมีราคาแพงกว่า STATUS ได้รับการทดสอบประสิทธิภาพแล้ว ในสถาบันต่างๆ เช่น British Antarctic Survey (BAS) และ British Architectural Library โปรแกรม STATUS มีความสามารถในการควบคุมฐานข้อมูลขนาดใหญ่ได้ไม่มีข้อจำกัดของขนาดระเบียบ โครงสร้างข้อมูลมีความยืดหยุ่น และโปรแกรมยังสามารถใช้ได้กับเครื่องคอมพิวเตอร์รุ่นต่างๆ ที่ได้มาตรฐานได้มาก

ส่วนการค้นคืนโปรแกรม STATUS สามารถค้นคืนได้โดยอิสระ โดยใช้กลยุทธ์การค้นคืนในลักษณะตรรกแบบบูลีน มีศัพท์สัมพันธ์ และสามารถค้นคืนแบบ truncation นอกจากนี้รูปแบบของผลลัพธ์สามารถเรียกใช้รูปแบบที่สร้างไว้ล่วงหน้าได้ โปรแกรม STATUS ไม่มีความสามารถในการใช้เป็นโปรแกรมประมวลผลคำ (Word Processing) แต่สามารถเชื่อมโยงกับโปรแกรมอื่นที่เป็นโปรแกรมประมวลผลคำได้ (Cook, 1986)

SPINDEX I, II, III เป็นระบบอัตโนมัติสำหรับการจัดการข้อมูลเกี่ยวกับต้นฉบับตัวเขียน และจดหมายเหตุ SPINDEX I เกิดขึ้นในหอสมุดรัฐสภาอเมริกันในปี ค.ศ. 1966 ออกแบบขึ้นมาเพื่อใช้ในการกรอกข้อมูลของ NUCMC (National union catalog of manuscript collections) data sheet แต่ไม่ได้รับการสนับสนุนจากหอสมุดรัฐสภาอเมริกันให้พัฒนาต่อไป จน Frank G. Burke ผู้พัฒนา SPINDEX I ได้ย้ายไปปฏิบัติงานในหอจดหมายเหตุแห่งชาติในปี ค.ศ. 1967 และเป็นกรรมการบริหารของ National Historical Publication and Records Commission (NHPRC) จึงได้พัฒนา SPINDEX II ขึ้นอีก เพื่อให้สามารถ

ควบคุมข้อมูลที่มีลักษณะ เป็นจดหมายเหตุต้นฉบับตัวเขียน และสามารถจัดทำรายการลงทะเบียน หรือทำคู่มือช่วยค้นจากดัชนีของ โปรแกรมได้ เป้าหมายของ โปรแกรม SPINDEX II คือ เพื่อทำการจัดเรียงข้อมูลที่ได้จากเอกสารต่างๆ เหล่านั้นจัดทำเป็น Permuted index ของเอกสาร

โปรแกรม SPINDEX III ได้รับการปรับปรุงจากโปรแกรม SPINDEX II ให้มีประสิทธิภาพมากขึ้น ทั้งในการทำรายการเอกสารที่ลงทะเบียน และการทำรายการดัชนี การประมวลผลของฐานข้อมูล SPINDEX II ต้องทำการประมวลผลทั้งแฟ้มข้อมูล แต่ SPINDEX III สามารถเลือกจำกัดขอบเขตของแฟ้มข้อมูลในการประมวลผลได้

ข้อจำกัดของ SPINDEX คือ การค้นคืนไม่มีลักษณะเป็นออนไลน์ เนื่องจากขาดมาตรฐานของระบบการค้นคืนข้อมูล จึงทำให้ต้องเลือกใช้การประมวลผลโดยระบบแบทช์ แทนที่จะเป็นลักษณะออนไลน์ การบันทึกข้อมูลต้องใช้ตามโครงสร้างที่ SPINDEX กำหนดให้ คู่มือที่ระบบผลิตให้จะเป็นรูปแบบที่กำหนดไว้ตามโปรแกรม คือ คู่มือการจัดเก็บ(Repository guide) และรายการเอกสารที่มี(Container list) (Berner, 1983)

GRIPHOS หรือ General Retrieval and Information Processing for Humanities-Oriented Studies พัฒนาโดยมหาวิทยาลัยแห่งรัฐนิวยอร์ก (State University of New York) และได้แพร่หลายไปยังพิพิธภัณฑ์ สถานที่แสดงศิลปะ เพื่อใช้ในการจัดหมวดหมู่ และควบคุมทรัพยากรประเภทภาพถ่าย โปรแกรมเขียนด้วยภาษา PL/1 และปฏิบัติงานบนเครื่อง และอุปกรณ์คอมพิวเตอร์ของ IBM โดยได้รับเงินสนับสนุนจาก (US) Museum Computer Network

OMSS เป็นระบบการค้นคืนสารนิเทศออนไลน์ที่พัฒนาขึ้น เพื่อใช้ในงานสารนิเทศ พิเศษของห้องสมุดของ Department of Hunter ของมหาวิทยาลัยเวสเทิร์น แคลิฟอร์เนีย (Western California University) โปรแกรมเขียนด้วยภาษาเบสิก บนเครื่อง VAX 11/780 (Cook, 1986)

จากการศึกษาการใช้คอมพิวเตอร์ในงานจดหมายเหตุมหาวิทยาลัย ทั้งในประเทศไทย และในต่างประเทศพบว่า ในประเทศไทยยังไม่มีการจัดทำฐานข้อมูลด้านจดหมายเหตุมหาวิทยาลัย

แต่ Somsak Vichien(1984) ได้ทดลองใช้โปรแกรมสำเร็จรูป CDS/ISIS บนเครื่องเมนเฟรม ในการออกแบบและผลิตดัชนีสำหรับเอกสารจดหมายเหตุของธนาคารแห่งประเทศไทย จำนวน 200 รายการ ซึ่งผู้วิจัยได้ศึกษาวิธีการดำเนินการสร้างฐานข้อมูล โดยเลือกเขตข้อมูลที่จำเป็น สำหรับการบันทึกข้อมูลจากกลุ่มตัวอย่าง และได้ใช้เป็นแนวความคิดว่า จะคัดเลือกเขตข้อมูลต่างๆ จาก US MARC: AMC เท่าที่จำเป็นสำหรับการบันทึกข้อมูลวัสดุจดหมายเหตุมหาวิทยาลัยศิลปากร ส่วนการศึกษาการใช้คอมพิวเตอร์ ในการดำเนินงานจดหมายเหตุมหาวิทยาลัย ในต่างประเทศทำให้ทราบว่า ในมหาวิทยาลัยต่างๆ ในต่างประเทศ มีการจัดเก็บและค้นคืน จดหมายเหตุมหาวิทยาลัยด้วยเครื่องคอมพิวเตอร์มาตั้งแต่ประมาณปี ค.ศ. 1977 โดยมีการใช้ ทั้งโปรแกรมสำเร็จรูป และโปรแกรมที่เขียนขึ้นด้วยภาษาคอมพิวเตอร์ วิธีการจัดเก็บส่วนใหญ่ จัดเก็บเป็นฐานข้อมูลเฉพาะจดหมายเหตุของมหาวิทยาลัย และสามารถค้นคืนได้ด้วยเทคนิคต่างๆ ในระบบเครือข่ายด้วย

การจัดเก็บและการค้นคืนสารนิเทศ

การจัดเก็บสารนิเทศ เป็นขั้นตอนของการป้อนข้อมูล(Input) ของเอกสาร สิ่งพิมพ์ ต่างๆ ลงระบบคอมพิวเตอร์ตามโปรแกรมที่วางไว้ ข้อมูลหรือเอกสารต่างๆ จะต้องผ่านการ วิเคราะห์เนื้อหาเพื่อให้คำสำคัญ หรือหัวเรื่อง และให้สาระสังเขป เพื่อใช้เป็นคำดัชนีในการ ค้นคืนข้อมูลจากระบบคอมพิวเตอร์ การจัดเก็บสารนิเทศจำนวนมากๆ ทำให้เกิดเป็นฐานข้อมูล ซึ่งเป็นแหล่งในการรวบรวมเอกสารสารนิเทศด้านต่างๆ และรอให้มีผู้มาแสดงความต้องการ ข้อมูลในฐานข้อมูลนั้น

การค้นคืนสารนิเทศเป็นการกระทำใดๆ ที่ทำให้ได้รับสารนิเทศที่ต้องการ ซึ่งอาจเป็น ข้อมูล หรือรายการเอกสารซึ่งบรรจุเนื้อหาที่ต้องการ การค้นคืนสารนิเทศมีความหมายเดียวกับ การค้นวรรณกรรม(literature searching) ซึ่งหมายถึงกระบวนการค้นเอกสารเพื่อให้ได้ เอกสารที่เกี่ยวข้องกับเรื่องใดเรื่องหนึ่ง หลักการสำคัญของการค้นคืนสารนิเทศ คือ การค้นหา และนำสารนิเทศที่ตรงตามความต้องการส่งให้แก่ผู้ใช้อย่างรวดเร็วทันการณ์(ชลทิชา สุทธิรัตนกุล, 2534: 2) Lancaster และ Owen (1976: 1) ให้ความหมายของการค้นคืนสารนิเทศว่า มักใช้เพื่อหมายถึงระบบ และขั้นตอนในการค้นคืนข้อมูลเอกสารด้านใดด้านหนึ่งที่ผู้ใช้ต้องการ โดยผู้ทำการค้นคืนจะวิเคราะห์คำถามของผู้ใช้บริการ และสร้างกลยุทธ์การค้นคืนให้สัมพันธ์

และเหมาะสมกับโปรแกรมคอมพิวเตอร์ที่ใช้อยู่ และกับ Input ของข้อมูล เช่น คำสำคัญ หรือ สารสนเทศของเอกสารในฐานะข้อมูล ผลลัพธ์ (Output) ที่ได้จะเป็นชุดของเอกสารที่คาดว่าจะมีความสัมพันธ์กับคำถามของผู้ใช้บริการ ส่วน Stokes (1982: 114) ให้ความหมายสั้นๆ ว่า คือ วิธีการทางระบบคอมพิวเตอร์ ที่ทำการค้นคืนรายการเอกสารจากแหล่งสะสมข้อมูล เช่น ฐานข้อมูล เพื่อให้ได้รายการเอกสารที่ต้องการ

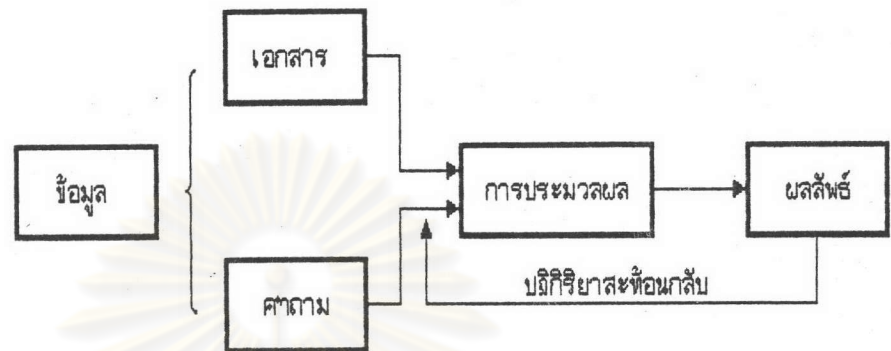
เมื่อรวมขบวนการในการจัดเก็บและค้นคืนสารนิเทศเข้าด้วยกันแล้ว จึงหมายความว่าถึง ขบวนการต่างๆ ในการรวบรวมข้อมูลเอกสารต่างๆ ไว้เป็นฐานข้อมูล และมีผู้ใช้บริการแสดงความต้องการข้อมูลจากเอกสารที่สะสมไว้ ตลอดจนขบวนการในการสร้างกลยุทธ์การค้นคืน ไม่ว่าจะทำการค้นคืนด้วยตนเอง หรือมีผู้ค้นคืนให้ โดยมีจุดประสงค์เพื่อพยายามให้ได้เอกสาร ที่ผู้ใช้บริการพึงประสงค์มากที่สุด (Rijsbergen, 1981: 5)

องค์ประกอบของระบบการจัดเก็บและค้นคืนสารนิเทศมีอยู่ 3 ส่วน คือ (Rijsbergen, 1981: 7)

1. การป้อนข้อมูล (Input)
2. การประมวลผล (Processing)
3. ผลลัพธ์ (Output)

และสามารถแสดงเป็นภาพได้ดังภาพประกอบที่ 2

ศูนย์วิจัยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



ภาพประกอบที่ 2 : องค์ประกอบของระบบการจัดเก็บและค้นคืนสารนิเทศ
(ดัดแปลงมาจาก A typical IR System ใน Rijsbergen 1981: 7)

1. การป้อนข้อมูล(Input) จะประกอบไปด้วยเอกสารที่รวมเป็นฐานข้อมูล และคำถามจากผู้ให้บริการ ที่มาแสดงความต้องการข้อมูลที่อยู่ในฐานข้อมูลนั้น โดยการตั้งคำถามเข้าสู่ระบบการจัดเก็บและค้นคืน
2. การประมวลผล(Processing) เป็นขบวนการในการทำงานของระบบคอมพิวเตอร์ เพื่อดึงเอารายการที่ผู้ต้องการออกมาจากฐานข้อมูล
3. ผลลัพธ์(Output) จะเป็นชุดของเอกสารที่ค้นคืนได้ และให้ผู้ให้บริการพิจารณาว่าตรงตามความต้องการมากหรือน้อยเพียงใด ซึ่งการพิจารณาของผู้ให้บริการจะส่งผลให้เป็นปฏิกิริยาย้อนกลับ(Feedback) เข้าสู่การป้อนคำถามลงสู่ระบบคอมพิวเตอร์อีกครั้งหนึ่ง เพื่อปรับปรุงกลยุทธ์การค้นคืน ในกรณีที่ได้อาไรการเอกสารที่ต้องการไม่เพียงพอ หรือมีรายการเอกสารที่ไม่ต้องการออกมามากเกินไป

ลักษณะของระบบการจัดเก็บและค้นคืนสารนิเทศที่สมบูรณ์ ที่ผู้ให้บริการต้องการ คือ สามารถค้นคืนเอกสารที่มีเนื้อหาตามที่ต้องการได้ทั้งหมดเท่าที่มีอยู่ในฐานข้อมูล โดยไม่มีรายการที่ไม่ต้องการติดมาด้วย แต่ในทางปฏิบัตินั้นผู้ต้องการข้อมูลจะต้องตรวจดูรายการที่มีอยู่ในฐานข้อมูล

ทั้งหมด จึงจะแน่ใจได้ว่าชุดของเอกสารที่ได้รับนั้นเป็นรายการที่ต้องการทั้งหมดที่มีในฐานข้อมูล ซึ่งจะสิ้นเปลืองเวลามาก หรือไม่สามารถทำได้เลยในกรณีที่เป็นฐานข้อมูลใหญ่ แต่หากว่าจะพยายามไม่ถึงเอารายการที่ไม่ต้องการออกมานั้น อาจจะทำได้โดยการกำหนดขอบเขตของข้อมูลที่ต้องการจากการใช้กลยุทธ์การค้นคืนแบบต่างๆ (Sharp, 1965: 14) นอกจากนี้ Kent (1966: 313) ยังให้ความเห็นว่า ระบบการค้นคืนสารนิเทศที่ดีจะต้องสามารถค้นคืนสารนิเทศที่ผู้ใช้ต้องการได้อย่างแม่นยำ แต่อย่างไรก็ตามก็ยังไม่สามารถรับรองได้ว่า ข้อมูลที่ถูกค้นคืนออกมาได้นั้นจะเป็นข้อมูลที่ผู้ใช้ต้องการใช้อย่างแท้จริง ดังนั้นจึงต้องมีการประเมินผลการค้นคืนสารนิเทศ เพื่อให้เข้าใจถึงสิ่งที่ผู้ใช้บริการต้องการอย่างแท้จริง

การประเมินผลระบบ

ระดับของการประเมินผลและการวัดคุณค่าระบบการบริการ Lancaster (1977: 1-2) แบ่งไว้ว่ามี 6 ระดับคือ

1. การวัดประสิทธิภาพ (Effectiveness) เป็นการวัดความพึงพอใจของผู้ใช้ที่มีต่อระบบด้วยการสำรวจความคิดเห็น โดยใช้แบบสอบถาม การสัมภาษณ์ หรือวัดความสำเร็จจากผลลัพธ์ที่ได้
2. การวัดประสิทธิภาพในด้านมูลค่าของระบบ (System's cost effectiveness) เป็นการวัดประสิทธิภาพการทำงานของระบบว่า ทำงานได้ตรงตามวัตถุประสงค์ และได้ผลลัพธ์ตรงตามความต้องการของผู้ใช้หรือไม่
3. การวัดความคุ้มค่า (Cost benefit) เป็นวิธีการวัดที่ทำได้ยาก เนื่องจากจะเกี่ยวกับคุณค่าของบริการที่ได้ โดยเป็นการวัดคุณค่า (Value) ของระบบบริการว่ามีมากหรือน้อยกว่าการลงทุนหรือไม่ คือ พยายามพิจารณาถึงค่าใช้จ่าย ที่ต้องใช้ในการให้บริการ ว่าเหมาะสมกับผลประโยชน์ที่ได้รับจากระบบหรือไม่
4. การวัดความคุ้มค่าในด้านประสิทธิภาพการทำงาน (Cost-performance benefit) เป็นการวัดความสัมพันธ์ระหว่างมูลค่าของการลงทุน (Cost) การทำงานของระบบ (Performance) และผลประโยชน์ (Benefit)
5. การวัดประสิทธิภาพในแบบ Macroevaluation เป็นการวัดการทำงานโดยรวมของระบบว่าสามารถทำงานได้ดีเพียงใด ผลของการวัดจะออกมาในรูปของระบบจำนวน

เช่น ร้อยละ

6. การวัดประสิทธิภาพในแบบ Microevaluation เป็นการหาเหตุผลของระดับในการทำงานของระบบ ว่าทำไมระบบจึงสามารถทำงานได้ในระดับนั้น

Salton และ McGill(1983: 158) สรุปว่าในการประเมินผลการค้นคืนสารนิเทศนั้น มีเหตุผลหลายประการเช่น

1. ต้องการเปรียบเทียบระบบที่มีอยู่เดิมกับระบบอื่นที่อาจต้องการนำมาใช้แทน
2. ต้องการพิจารณาว่าประสิทธิภาพของระบบเปลี่ยนไปอย่างไร ถ้ามีส่วนประกอบใดส่วนประกอบหนึ่งของระบบเปลี่ยนแปลงไป เช่น พิจารณาว่าประสิทธิภาพของระบบการค้นคืนสารนิเทศได้เปลี่ยนแปลงไป เมื่อประเภทของคำถามเปลี่ยนแปลงหรือหัวเรื่องเปลี่ยนแปลง
3. เพื่อประเมินผลส่วนประกอบใหม่ของระบบที่คาดว่าจะนำไปรวมกับระบบที่มีอยู่เดิม ซึ่งถ้าเป็นในกรณีนี้ การทำงานของระบบใหม่ควรมีการนำไปใช้ในลักษณะร่วมกันไปก่อนก่อนที่จะมีการสร้างระบบใหม่ที่แท้จริงขึ้น

Pao (1989: 215) กล่าวถึงเหตุผลในการประเมินผลระบบการค้นคืนสารนิเทศว่า มักจะตั้งขึ้นจากวัตถุประสงค์ในการประเมินผลว่ามีวัตถุประสงค์อย่างไร เช่น เพื่อพิจารณาถึงความคุ้มค่าในการลงทุนติดตั้งระบบ หรือประเมินค่าใช้จ่ายของระบบ ซึ่ง Pao ให้ความเห็นว่า เหตุผลในการประเมินผลระบบการค้นคืนเพื่อพิจารณาถึงความคุ้มค่านั้นมี 4 ประการคือ

1. เพื่อชี้แจงรายการในการลงทุน สำหรับเป็นข้อมูลพิจารณาค่าในการทำงานของระบบ และนำเสนอค่าใช้จ่ายเพื่อให้ผู้บริหารสำรวจคุณภาพ และปริมาณในการทำงานของระบบ
2. เพื่อก่อให้เกิดความเชื่อมั่นในคุณภาพของการบริการ โดยประเมินผลว่าระบบสามารถทำงานได้บรรลุวัตถุประสงค์ได้ดีเพียงใด
3. เพื่อให้เกิดการวิจารณ์ในการแก้ปัญหาของระบบการค้นคืนสารนิเทศที่เกิดขึ้น
4. เพื่อทำนายคุณภาพในการทำงานของระบบในอนาคตได้อย่างมีเหตุผล

ส่วน Salton และ McGill (1983: 158) แบ่งระดับการประเมินผลออกเป็น 2 ประเภท คือ

1. ประเมินประสิทธิภาพของระบบ (System Effectiveness) หมายถึง การประเมินผลความสามารถในการให้บริการสารสนเทศตามความต้องการของผู้ใช้บริการ
2. ประเมินประสิทธิภาพของการทำงานของระบบ (Efficiency of the Operations) เป็นการวัดคุณค่า ความคุ้มค่า หรือเวลาที่ต้องสูญเสียไปในการปฏิบัติงานนั้นๆ

ตัวแปรในการประเมินผลระบบการค้นคืนสารสนเทศ

ตัวแปรที่มีผลต่อความพึงพอใจของผู้ใช้บริการมี 6 ประการ คือ (Cleverdon, 1964 quoted in Salton, 1968: 282)

1. ความครอบคลุมของเนื้อหาของเอกสารในฐานะข้อมูล (Coverage)
2. เวลาเฉลี่ยของระยะเวลาที่ใช้ในการค้นคืนจนได้รับคำตอบออกมา (Time lag)
3. รูปแบบของการเสนอผลลัพธ์ที่ได้ (Presentation)
4. ความพยายามที่ใช้ไปในส่วนที่จะให้ผู้บริการได้คำตอบ (Effort)
5. ค่า Recall ระบบ
6. ค่า Precision ของระบบ

ซึ่ง Cleverdon กล่าวว่า ประเด็นที่ 1-4 นั้น สามารถประเมินผลได้ง่าย และถ้ามีส่วนใดส่วนหนึ่งที่ทำให้ผู้ใช้ไม่พึงพอใจ ก็สามารถปรับปรุงได้โดยดัดแปลงแก้ไขลักษณะของระบบ แต่ตัวแปรอีก 2 ประเด็น คือ ค่า Recall และค่า Precision นั้นประเมินผลได้ยาก และต้องใช้ความพยายามในการปรับปรุงประสิทธิภาพถ้าหากผู้ใช้ไม่พึงพอใจ บางครั้งมีการโต้แย้งว่าค่า Recall และค่า Precision นั้นจะสามารถใช้วัดประสิทธิภาพได้อย่างไร เพราะค่า Recall และค่า Precision นั้นเป็นเพียงค่าโดยประมาณ อย่างไรก็ตาม Salton และ McGill (1983: 172), Kiewitt (1979: 88) และ Kochen (1974: 156) มีความเห็นว่าการนำค่า Recall และค่า Precision มาใช้วัดประสิทธิภาพของระบบการค้นคืนนั้น เป็นวิธีหลักที่มีผู้นิยมใช้มากที่สุด และเป็นที่ยอมรับว่าสามารถใช้ประเมินผลได้ดี และสามารถสนองตอบ

ความต้องการของผู้ใช้บริการได้มาก นอกจากนี้ยังมีลักษณะเป็นเลขจำนวนซึ่งสามารถเข้าใจได้ง่าย

วิธีการประเมินผลการค้นคืนสารนิเทศโดยใช้ตัวแปร Recall และ Precision นั้น นักทฤษฎีทางด้านสารนิเทศมีวิธีการคิด และมีสูตรที่ใช้คำนวณค่า Recall และค่า Precision ที่คล้ายคลึงกัน เพียงแต่มีการใช้พารามิเตอร์ ที่ใช้แทนความหมายของส่วนของเอกสารที่ต้องการ กล่าวถึงต่างกัน โดยในการประเมินผลการค้นคืนสารนิเทศโดยใช้ตัวแปร Recall และ Precision จะต้องมีส่วนประกอบของเอกสารที่ใช้ในการคำนวณ 4 ส่วนคือ (Jones, 1971: 95)

1. เอกสารที่ค้นคืนออกมาได้และสัมพันธ์กับเรื่องที่ต้องการ
2. เอกสารที่ค้นคืนออกมาได้แต่ไม่สัมพันธ์กับเรื่องที่ต้องการ
3. เอกสารที่สัมพันธ์กับเรื่องที่ต้องการแต่ไม่ถูกค้นคืนออกมา
4. เอกสารที่ไม่สัมพันธ์กับเรื่องที่ต้องการและไม่ถูกค้นคืนออกมา

เช่นเดียวกับ Salton (1968: 283-284) และ Fairthorne (Fairthorne, quoted in Farradane, 1978: 195-209) ให้ความเห็นว่าถ้าจะใช้ค่า Recall และค่า Precision วัดประสิทธิภาพของการค้นคืน จะต้องแบ่งเอกสารทั้งหมดที่มีอยู่ในฐานข้อมูลออกเป็น 4 ส่วน โดยมีการแบ่งออกเป็น 2 ด้าน คือด้านแรกเป็นการแบ่งในส่วนที่เกี่ยวข้องกับการพิจารณาของผู้ใช้บริการ (User) คือ แบ่งเป็นเอกสารที่สัมพันธ์กับเรื่อง que ผู้ใช้บริการต้องการ และเอกสารที่ไม่สัมพันธ์กับเรื่อง que ผู้ใช้บริการต้องการ ด้านที่สองเป็นการแบ่งในส่วนที่เกี่ยวข้องกับระบบ (System) คือแบ่งเป็นเอกสารที่ค้นคืนได้ และเอกสารที่ค้นคืนออกมาไม่ได้ โดย Fairthorne อธิบายว่า เมื่อมีผู้ค้นคืนข้อมูลจากฐานข้อมูล และระบบคอมพิวเตอร์ทำงานตามคำสั่งแล้ว ผลลัพธ์ (Output) ที่ระบบคอมพิวเตอร์ค้นคืนได้ จะเป็นข้อมูลที่สัมพันธ์ หรือคาดว่าสัมพันธ์กับคำถามที่ผู้ค้นคืนป้อนความต้องการลงสู่ระบบ แต่อย่างไรก็ตามอาจจะยังมีเอกสารบางส่วนที่ไม่สามารถค้นคืนออกมาได้ (Retrieved/Not Retrieved) ต่อจากนั้นผู้ใช้บริการหรือผู้ที่ต้องการข้อมูล จะเป็นผู้พิจารณาว่า ชุดของข้อมูลที่ได้นั้นมีความสัมพันธ์กับคำถามหรือไม่ (Relevant/Not Relevant)

รูปแบบพารามิเตอร์ และสูตรที่ใช้ในการคำนวณค่าRecall และค่าPrecision ที่นัก
ทฤษฎีทางด้านสารสนเทศใช้ ได้รวบรวมไว้ในตารางที่ 2



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 2 : รูปแบบ พารามิเตอร์ ความหมาย และสูตรที่ใช้ในการคำนวณ Recall และ Precision

นักทฤษฎี	รูปแบบที่ใช้	พารามิเตอร์ที่ใช้ และความหมาย	สูตรที่ใช้ในการคำนวณ																
Cleverdon*	Contingency Table	<table border="1"> <thead> <tr> <th></th> <th>Relevant</th> <th>Not Relevant</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>Retrieved</th> <td>Hits (a)</td> <td>False drops (b)</td> <td></td> </tr> <tr> <th>Not Retrieved</th> <td>Misses (c)</td> <td>Dodged (d)</td> <td></td> </tr> <tr> <th>Total</th> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>Hits คือ เอกสารที่ค้นคืนออกมาได้และสัมพันธ์กับเรื่องที่ต้องการ False drops คือ เอกสารที่ค้นคืนออกมาได้แต่ไม่สัมพันธ์กับเรื่องที่ต้องการ Misses คือ เอกสารที่สัมพันธ์กับเรื่องที่ต้องการแต่ไม่ถูกค้นคืนออกมา Dodged คือ เอกสารที่ไม่สัมพันธ์กับเรื่องที่ต้องการและไม่ถูกค้นคืนออกมา</p>		Relevant	Not Relevant	Total	Retrieved	Hits (a)	False drops (b)		Not Retrieved	Misses (c)	Dodged (d)		Total				$\text{Recall} = \frac{a}{a + c}$ $\text{Precision} = \frac{a}{a + b}$
	Relevant	Not Relevant	Total																
Retrieved	Hits (a)	False drops (b)																	
Not Retrieved	Misses (c)	Dodged (d)																	
Total																			


* Cleverdon, quoted in Pao, 1989: 226-227

ตารางที่ 2 : รูปแบบ พารามิเตอร์ ความหมาย และสูตรที่ใช้ในการคำนวณค่า Recall และ Precision(ต่อ)

นักทฤษฎี	รูปแบบที่ใช้	พารามิเตอร์ที่ใช้ และความหมาย	สูตรที่ใช้ในการคำนวณ												
Fairthorne*	Contingency Table	<p style="text-align: center;">Not Retrieved Retrieved Total</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="text-align: center;">Relevant</td> <td style="text-align: center;">R</td> <td style="text-align: center;">C-R</td> <td style="text-align: center;">C</td> </tr> <tr> <td style="text-align: center;">Not Relevant</td> <td style="text-align: center;">L-R</td> <td style="text-align: center;">N-L-C+R</td> <td style="text-align: center;">N-C</td> </tr> <tr> <td style="text-align: center;">Totals</td> <td style="text-align: center;">L</td> <td style="text-align: center;">N-L</td> <td style="text-align: center;">N</td> </tr> </table> <p>R คือ เอกสารที่ค้นคืนออกมาได้และสัมพันธ์กับเรื่องที่ต้องการ L-R คือ เอกสารที่ค้นคืนออกมาได้แต่ไม่สัมพันธ์กับเรื่องที่ต้องการ C-R คือ เอกสารที่สัมพันธ์กับเรื่องที่ต้องการแต่ไม่ถูก ค้นคืนออกมา N-L-C+R คือ เอกสารที่ไม่สัมพันธ์กับเรื่องที่ต้องการและไม่ถูกค้นคืนออกมา</p>	Relevant	R	C-R	C	Not Relevant	L-R	N-L-C+R	N-C	Totals	L	N-L	N	$\text{Recall} = \frac{R}{L}$ $\text{Precision} = \frac{R}{C}$
Relevant	R	C-R	C												
Not Relevant	L-R	N-L-C+R	N-C												
Totals	L	N-L	N												

* Fairthorne, 1964 quoted in Farradane, 1974: 202

ตารางที่ 2 : รูปแบบ พารามิเตอร์ ความหมาย และสูตรที่ใช้ในการคำนวณค่า Recall และ Precision(ต่อ)

นักทฤษฎี	รูปแบบที่ใช้	พารามิเตอร์ที่ใช้ และความหมาย	สูตรที่ใช้ในการคำนวณ						
Swets*	Contingency Table	<p style="text-align: center;">Not Retrieved Retrieve</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding: 5px;">Relevant</td> <td style="padding: 5px; text-align: center;">Hit</td> <td style="padding: 5px; text-align: center;">Miss</td> </tr> <tr> <td style="padding: 5px;">Irrelevant</td> <td style="padding: 5px; text-align: center;">Trash</td> <td style="padding: 5px; text-align: center;">Pass</td> </tr> </table> <p>Hit คือ เอกสารที่ค้นคืนออกมาได้ และสัมพันธ์กับเรื่องที่ต้องการ</p> <p>Miss คือ เอกสารที่ค้นคืนออกมาได้ แต่ไม่สัมพันธ์กับเรื่องที่ต้องการ</p> <p>Trash คือ เอกสารที่สัมพันธ์กับเรื่องที่ต้องการ แต่ไม่ถูกค้นคืนออกมา</p> <p>Pass คือ เอกสารที่ไม่สัมพันธ์กับเรื่องที่ต้องการและไม่ถูกค้นคืนออกมา</p>	Relevant	Hit	Miss	Irrelevant	Trash	Pass	$h (\text{Recall}) = \frac{\text{Hit}}{\text{Hit} + \text{Miss}}$ $a (\text{Precision}) = \frac{\text{Hit}}{\text{Hit} + \text{Trash}}$ <p>$h = \text{hit rate}$</p> <p>$a = \text{acceptance rate}$</p> <div style="text-align: right;">  </div>
Relevant	Hit	Miss							
Irrelevant	Trash	Pass							

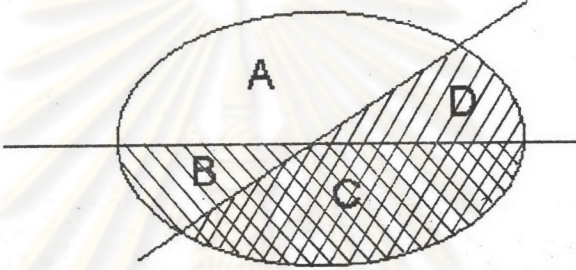
* Swets, 1963 quoted in Kochen 1974: 157

ตารางที่ 2 : รูปแบบ พารามิเตอร์ ความหมาย และสูตรที่ใช้ในการคำนวณหาค่า Recall และ Precision(ต่อ)

นักทฤษฎี	รูปแบบที่ใช้	พารามิเตอร์ที่ใช้ และความหมาย	สูตรที่ใช้ในการคำนวณ												
Rijsbergen*	Contingency Table	<p>Ret Not Rel Rel</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding: 5px;">Ret</td> <td style="padding: 5px;">$A \cap B$</td> <td style="padding: 5px;">$\bar{A} \cap B$</td> <td style="padding: 5px;">B</td> </tr> <tr> <td style="padding: 5px;">Not Ret</td> <td style="padding: 5px;">$A \cap \bar{B}$</td> <td style="padding: 5px;">$\bar{A} \cap \bar{B}$</td> <td style="padding: 5px;">\bar{B}</td> </tr> <tr> <td></td> <td style="padding: 5px;">A</td> <td style="padding: 5px;">\bar{A}</td> <td style="padding: 5px;">N</td> </tr> </table> <p>$A \cap B$ คือ เอกสารที่ค้นคืนออกมาได้และสัมพันธ์กับเรื่องที่ต้องการ</p> <p>$\bar{A} \cap B$ คือ เอกสารที่ค้นคืนออกมาได้แต่ไม่สัมพันธ์กับเรื่องที่ต้องการ</p> <p>$A \cap \bar{B}$ คือ เอกสารที่สัมพันธ์กับเรื่องที่ต้องการ แต่ไม่ถูกค้นคืนออกมา</p> <p>$\bar{A} \cap \bar{B}$ คือ เอกสารที่ไม่สัมพันธ์กับเรื่องที่ต้องการ และไม่ถูกค้นคืนออกมา</p>	Ret	$A \cap B$	$\bar{A} \cap B$	B	Not Ret	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	\bar{B}		A	\bar{A}	N	$\text{Recall} = \frac{ A \cap B }{ A }$ $\text{Precision} = \frac{ A \cap B }{ B }$
Ret	$A \cap B$	$\bar{A} \cap B$	B												
Not Ret	$A \cap \bar{B}$	$\bar{A} \cap \bar{B}$	\bar{B}												
	A	\bar{A}	N												

* Rijsbergen 1981: 148-149

ตารางที่ 2 : รูปแบบ พารามิเตอร์ ความหมาย และสูตรที่ใช้ในการคำนวณค่า Recall และ Precision(ต่อ)

นักทฤษฎี	รูปแบบที่ใช้	พารามิเตอร์ที่ใช้ และความหมาย	สูตรที่ใช้ในการคำนวณ
Salton*	รูปภาพ	 <p>// Retrieved \\ Relevant</p> <p>A คือ เอกสารที่ค้นคืนออกมาได้และสัมพันธ์กับเรื่องที่ต้องการ</p> <p>B คือ เอกสารที่ค้นคืนออกมาได้แต่ไม่สัมพันธ์กับเรื่องที่ต้องการ</p> <p>C คือ เอกสารที่สัมพันธ์กับเรื่องที่ต้องการ แต่ไม่ถูกค้นคืนออกมา</p> <p>D คือ เอกสารที่ไม่สัมพันธ์กับเรื่องที่ต้องการ และไม่ถูกค้นคืนออกมา</p>	$\text{Recall} = \frac{c}{b + c}$ $\text{Precision} = \frac{c}{c + d}$

* Salton, 1968: 284

จากการใช้ตาราง Contingency หรือ รูปภาพ แสดงการแบ่งเอกสารในฐานข้อมูล เพื่อใช้ในการคำนวณค่า Recall และค่า Precision และวิธีการคำนวณค่า Recall และค่า Precision ของนักทฤษฎีทางด้านสารสนเทศแต่ละคนแล้ว สามารถสรุปเป็นสูตรที่ใช้ในการคำนวณค่า Recall และค่า Precision โดยคิดเป็นค่าร้อยละได้ดังนี้

$$\text{Recall} = \frac{\text{จำนวนของเอกสารที่ค้นคืนได้และสัมพันธ์กับเรื่องที่ต้องการ}}{\text{จำนวนของเอกสารทั้งหมดที่สัมพันธ์กับเรื่องที่ต้องการที่มีในฐานข้อมูล}} \times 100$$

$$\text{Precision} = \frac{\text{จำนวนของเอกสารที่ค้นคืนได้และสัมพันธ์กับเรื่องที่ต้องการ}}{\text{จำนวนของเอกสารที่ถูกค้นคืนออกมาทั้งหมด}} \times 100$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย