

# การสร้างบัญชีคำ (Thesaurus Construction)

จิรวรรณ ภักดีบุตร\*

## วิวัฒนาการของ Thesaurus

ในช่วงปี ค.ศ. 1939-1948 เป็นระยะที่มีการคิดค้นหาวิธีการใหม่ ๆ ในการจัดทำบรรณานุกรมและวิธีการค้นหาเอกสารที่มีประสิทธิภาพยิ่งขึ้น เริ่มต้นด้วย W.E. Batten ได้คิดค้นระบบการจัดทำบรรณานุกรมค้นหาสิทธิบัตร (patents) และไล่เรียงกันนั้น ในประเทศฝรั่งเศส G. Cordonnier ได้คิดค้นวิธีการที่เรียกว่า optical coincidence หรือ peek-a-boo ขึ้น โดยมีแนวคิดเกี่ยวกับ Batten ต่อมาใน ค.ศ. 1947-1948 Calvin Mooers คิดค้น Lato coding system ขึ้น และเป็นครั้งแรกที่เรียกคำศัพท์ที่ใช้แสดงเนื้อหาเอกสารว่า "descriptors"

ประมาณ ค.ศ. 1951 Mortimer Taube ได้นำสิ่งใหม่เข้ามาในวงการ คือ ระบบการทำบรรณานุกรมและค้นหาข้อมูลที่เรียกว่า Uniterm system โดยใช้คำโดดคำเดียว (single word) คำศัพท์ที่ใช้จะนำมาจากคำที่ใช้ในเอกสารซึ่งเป็นภาษาธรรมชาติ (natural language) โดยไม่มีการควบคุมคำศัพท์แต่อย่างใด (uncontrolled vocabulary) การค้นหาข้อมูลจะกระทำโดยการนำเอาคำศัพท์ที่เป็นคำโดดเหล่านี้มาใส่ลงในบัตรพร้อมด้วยเลขทะเบียนเอกสาร

อย่างไรก็ตาม วิธีการให้คำศัพท์หรือ index terms ที่ไม่มีการควบคุมคำศัพท์ (uncontrolled vocabulary) ก่อให้เกิดปัญหาหลายประการ เช่น

1. เอกสารที่มีเนื้อหาเหมือนกัน อาจใช้ index terms ที่แตกต่างกันไปตามคำศัพท์ที่พบและใช้ในเอกสารแต่ละฉบับ
2. คำศัพท์หรือ index terms ที่เขียนเหมือนกันอาจมีความหมายแตกต่างกันได้ เช่น Boxers (ones who box) หรือ Boxers (dogs)

\* จิรวรรณ ภักดีบุตร Ph.D. สำนักบรรณสารการพัฒนา สถาบันบัณฑิตพัฒนบริหารศาสตร์

3. คำศัพท์หรือ index terms ให้ความหมายกว้าง ๆ ทำให้ต้องใช้เวลาในการค้นหา

4. ความสัมพันธ์ระหว่างคำศัพท์แต่ละคำอาจไม่ชัดเจน อาจก่อให้เกิด false coordination และ incorrect term relationship ได้

5. การตีความหมายของคำศัพท์ขึ้นอยู่กับผู้ทำตรรกษี เช่น wood อาจหมายถึง a product of the lumber industry, a raw material, หรือ a fuel ก็ได้

ต่อมาจึงได้นำเอาคำศัพท์ผสม bound terms หรือ multiword terms มาใช้ แต่เดิมมานั้น ระบบการทำตรรกษี และการค้นหาจะใช้ระบบ Pre-coordinate system ต่อมาจึงค่อย ๆ เป็น post-coordinate system

Pre-coordinate system คือระบบที่คำศัพท์ต่าง ๆ ไม่ว่าจะเป็น subject headings หรือ index terms ได้มีการกำหนดมาอย่างแน่นอนและในการค้นหาจะต้องใช้คำศัพท์เหล่านั้นค้น โดยไม่อาจนำมาผสมคำขึ้นใหม่ได้ บางแห่งเรียกระบบนี้ว่า Non-manipulative system

Post-coordinate system คือระบบที่คำศัพท์ index terms อาจเป็นคำโดดหรือคำผสมกำหนดไว้ และในการค้นหาจะยอมให้มีการ match terms กันเพื่อให้ได้ข้อมูลที่ต้องการ อาจใช้วิธีการ Boolean logic (AND, OR and NOT) บางแห่งเรียกระบบนี้ว่า manipulative system

ในระยะต้นได้มีการค้นหาแบบ pre-coordinate system ต่อมาจากการคิดค้นของบุคคลทั้ง 4 ข้างต้น โดยเฉพาะระบบ Uniterm ของ Taube ซึ่งได้ใช้ระบบกึ่ง post-coordinate system ขึ้น แต่ไม่สามารถจะแก้ปัญหาที่เกิดขึ้นกับระบบ Uniterm ได้ เนื่องจากระบบ Uniterm เป็นระบบที่ไม่มีการควบคุมรายการคำศัพท์ ต่อมาจึงใช้ระบบ vocabulary control ทำให้เกิดระบบ Thesaurus ขึ้น

“Thesaurus” เป็นคำที่ใช้กันครั้งแรกในประเทศอังกฤษ ประมาณ ค.ศ. 1956 และรู้จักกันมากขึ้นจากการตีพิมพ์ของ Hans Peter Luhn แห่งบริษัท IBM ใน ค.ศ. 1957 thesaurus ฉบับแรกที่ใช้เพื่อควบคุมบัญญัติคำศัพท์ในระบบสารนิเทศได้จัดทำโดยบริษัท Dupont ใน ค.ศ.

## โครงสร้างของ Thesaurus

Thesaurus คือรายการคำศัพท์ หรือบัญญัติคำศัพท์ ซึ่งประกอบด้วยคำศัพท์ (หรือหัวเรื่อง) ที่เรียกว่า descriptors ถ้าเป็นรายการคำศัพท์ในสาขาวิชาเดียวเรียกว่า micro-thesaurus ถ้ารวมหลายสาขาวิชาเข้าด้วยกัน เรียกว่า macro-thesaurus Thesaurus อาจเปรียบเทียบกับรายการหัวเรื่อง (List of Subject Headings)

โดยทั่วไปรายการหัวเรื่องจะประกอบด้วยคำศัพท์หรือหัวเรื่อง (subject headings) โดย

1. หัวเรื่องจะเรียงตามลำดับอักษรของหัวเรื่องที่ปรากฏทั้งหมด ก-ข หรือ A-Z
2. ถ้าเป็นรายการหัวเรื่องรวมหลายสาขาวิชา รายการหัวเรื่องนั้นจะไม่มีการจัดหมวดหมู่คำศัพท์ หรือแยกคำศัพท์ตามสาขาวิชา คำศัพท์หรือหัวเรื่องในสาขาวิชาเดียวกัน หรือที่มีความสัมพันธ์เกี่ยวเนื่องกันจะกระจายกันอยู่ในรายการหัวเรื่องทั้งหมดตามลำดับอักษร โดยมีรายการโยง (cross references) เช่น ดู, ดูเพิ่มเติม, see, see also, และ see also from โยงไปมาถึงกัน

ในระบบ post-coordinate ได้ใช้ thesaurus เป็นสิ่งที่ควบคุมคำศัพท์ (vocabulary control) thesaurus มีโครงสร้างและการจัดระบบคล้ายคลึงกับรายการหัวเรื่อง แต่วิธีการที่แสดงถึงความสัมพันธ์ระหว่างคำศัพท์ descriptors แตกต่างกันโดย thesaurus จะให้รายการคำศัพท์ตามลำดับอักษร ควบคุมคำที่มีความหมายเหมือนกัน คำพ้องหรือคำที่เขียนเหมือนกันแต่ความหมายแตกต่างกัน มีรายการแสดงถึงความสัมพันธ์ระหว่าง generic-specific และความสัมพันธ์อื่นอีกระหว่าง descriptors thesaurus โดยทั่วไปมีองค์ประกอบดังนี้

1. ทะเบียนคำที่จัดไว้ตามลำดับอักษร (Alphabetical Index หรือ Listing) ทะเบียนคำจะเป็น descriptors ที่ใช้ในสาขาวิชานั้นทั้งหมดจัดเรียงตามลำดับอักษร รวมคำที่กำหนดให้ใช้และไม่กำหนดให้ใช้รวมทั้งการแนะนำว่าคำใดให้ใช้ได้ และใช้แทนคำอะไรบ้างเช่น

Administrative Courts

WE 1805

UF Administrative Tribunals

BT Public Administration

RT Administrative Law

Courts

Law

## Administrative Law

WE 1908

BT Public Administration

NT Administrative Procedures

RT Administrative Courts  
Law

## Administrative Procedures

WE 2008

BT Administrative Law

RT Civil Law and Procedures  
Management

Public Administration

## Administrative Tribunals

Use : Administrative Courts

## Administrators and Managers

MK 7702

UF Executives

BT Occupation

NT Records Managers

RT Appointed Officials

Court Administration

Elected Officials

Management

Mayors

Public Administration

School Management

State Governors

U.S. Presidents

2. ทะเบียนคำที่จัดไว้ตามหัวข้อใหญ่เพื่อแสดงถึงความเกี่ยวเนื่องของสาขาวิชา

(Hierarchical Display) โดยแสดงถึงความเกี่ยวเนื่องตามลำดับชั้นของ descriptors แต่ละตัว

**Table 1 Overview of the hierarchy**

AA0109	PHYSICAL ENVIRONMENT
AB0108	CLIMATE
AE0105	NATURAL EVENTS
AH0102	NATURAL RESOURCES
AK0102	GEOGRAPHIC ENVIRONMENT
HA0108	DEVELOPMENT AND INFRASTRUCTURE
HB0107	DEVELOPMENTAL ACTIVITIES
HK0101	ARCHITECTURE
HM0109	CONSTRUCTION
HS0103	INFRASTRUCTURE
HS0206	COMMUNICATION AND INFORMATION SYSTEMS
HT0105	HOUSING
HW0102	SANITATION
HX0101	TRANSPORTATION
MA0102	SOCIAL ENVIRONMENT
MB0101	CULTURAL ACTIVITIES
ME0108	ECONOMIC ACTIVITY
MK0105	HUMAN RESOURCES
MM0103	POPULATION
MP0100	INFORMATION BASED PROCESSES
MS0107	SOCIAL PROCESSES AND EVENTS
MX0105	SOCIAL CONDITIONS
TA0100	PUBLIC AFFAIRS
TB0109	CIVIL RIGHTS
TE0106	CRIMINAL JUSTICE
TH0103	ENVIRONMENTAL PROTECTION
TL0102	HEALTH
TP0108	MANPOWER
TS0105	PLANNING
TW0104	POLITICAL AND CIVIC ACTIVITIES
TX0103	SOCIAL WELFARE
WA0101	SOCIAL INSTITUTIONS
WB0100	GOVERNMENTAL INSTITUTIONS
WE0107	LAW
WE1702	PUBLIC ADMINISTRATION
WH0104	EXECUTIVE BRANCH
WL0103	JUDICIARY
WM0102	LEGISLATIVE BODIES
WS0106	ECONOMIC INSTITUTIONS
WW0105	MANAGEMENT
WX0104	BUSINESS ORGANIZATION
XA0108	BUSINESS ENTERPRISES
XHQ101	EDUCATION
XL0100	FAMILY
XP0106	RELIGIOUS INSTITUTIONS
XT0105	ASSOCIATIONS

The Urban Information Thesaurus : A Vocabulary for Social Documentation

MA 0102 to MB 4305

BRANCH 3: SOCIAL ENVIRONMENT

MA 0102	SOCIAL ENVIRONMENT
MB 0101	Cultural Activities
MB 0204	Cultural Facilities
MB 0307	Archives
MB 0385	Historic Buildings and Sites
MB 0400	Historical Societies
MB 0503	Libraries
MB 0606	Museums
MB 0709	Theatres
MB 0802	Concert Halls
MB 0905	Arts
MB 1005	Fine Arts
MB 1108	Painting
MB 1201	Photography
MB 1304	Sculpture
MB 1407	Performing Arts
MB 1500	Dance
MB 1603	Drama
MB 1706	Music
MB 1809	Symphony Orchestras
MB 1902	Opera
MB 2002	Crafts
MB 2105	Fairs and Festivals
MB 2208	Gambling
MB 2301	Illegal Gambling
MB 2404	Legalized Gambling
MB 2507	Off Track Betting
MB 2600	Lotteries
MB 2703	Leisure Activities
MB 2806	Recreation
MB 2909	Recreational Facilities
MB 3009	Arenas and Stadia
MB 3102	Camps
MB 3205	Marinas
MB 3308	Playgrounds and Playfields
MB 3401	Parks
MB 3504	National Parks and Reserves
MB 3607	State Parks
MB 3700	Local Parks
MB 3803	Open Space Land
MB 3906	Public Lands
MB 4006	Race Tracks
MB 4109	Resorts
MB 4202	Trails
MB 4305	Zoological and Botanical Gardens

3. รหัสประจำคำศัพท์ (Notation หรือ Code number) เพื่อจัดระเบียบหัวเรื่องให้รวมกลุ่มกันภายใต้หมวดวิชาใหญ่ต่าง ๆ ได้ตามความเกี่ยวเนื่องตามลำดับชั้นของ descriptors ตามข้อ 2 ข้างต้น รหัสประจำคำศัพท์อาจใช้เป็นสัญลักษณ์การจัดหมวดหมู่และการจัดเรียงเอกสารบนชั้นตามหมวดวิชาใหญ่ก็ได้

4. ให้นิยาม หรือขอบเขตของคำศัพท์ที่ใช้ (Scope Note = SN) เพื่อเน้นความหมายของ descriptors แต่ละตัวโดยเฉพาะไม่ให้นำมาใช้ผิดพลาด เช่น

#### INCOME

SN Income of an individual person or organization.

otherwise use National Income

5. มีการแสดงให้เห็นถึงความเกี่ยวข้อง/ความสัมพันธ์ (Relationship) และความเกี่ยวเนื่องตามลำดับชั้น (Hierarchy) ของ descriptors โดยนำเอารายการโยง (cross referenees) มาใช้

#### Use

UF (Used For)

BT (Broader Term)

NT (Narrower Term)

และ SN (Scope Note)

TT (Top Term)

เช่น INCOME No 9.20 \* N 98

SN Income for an individual person or organization.

Otherwise use National

NT Capital gains

Educational Income

BT Financial Resources

TT Economics

Resources

RT Assets  
Expenditures  
.....

และ

TAX

Use

Taxation

Taxation

UF Tax

NT Consumption tax

Customs duties

Income tax

RT Fiscal policy

Thesaurus บางสาขาอาจใช้ Permuted Descriptor Display โดยนำเอาคำที่มีความสำคัญใน descriptors ไม่ว่าจะ เป็นคำโดดหรือคำผสมมาใช้เป็น entry word เช่น

Vocational Aptitude

Vocational Counseling  
.....

Adult Vocational Education

Vocational Education Teachers  
.....

Area Vocational Schools

Vocational Training Centers  
.....

Eye Voice Span

Voluntary Agencies  
.....





schools

Teachers

Universities and colleges

X Instruction

XX Civilization

Coeducation

Culture

Schools

## INSTRUCTION

See EDUCATION

ถ้านำตัวอย่างการให้หัวเรื่องข้างต้นมาใช้ใน thesaurus จะเป็นดังนี้

## EDUCATION

UF Instruction

NT Higher education

Primary education

Secondary education

Vocational education

BT Civilization

RT Colleges

Schools

Universities

## INSTRUCTION

Use EDUCATION

## วิธีการจัดทำ Thesaurus

การจัดทำ thesaurus (controlled vocabulary) มีกระบวนการดังนี้

1. กำหนดหัวข้อเรื่อง/สาขาวิชาที่จะจัดทำ

2. คัดเลือกคำศัพท์ descriptors ที่จะใช้รวมทั้งกำหนดรูปแบบที่จะใช้ รวบรวม descriptors และแสดงรายการ descriptors

วิธีการจัดทำ descriptors มีวิธีการ 4 แบบ คือ

1. จัดทำคำศัพท์ descriptors จากเนื้อหาสาระของเอกสารกลุ่มหนึ่งที่มีอยู่ในสาขาวิชานั้น โดยรวบรวมเป็น descriptors เพื่อใช้ในระบบการจัดและค้นหาข้อมูลของตนเอง
2. แก้ไขปรับปรุงจากรายการคำศัพท์ที่มีอยู่ เช่น จากรายการหัวเรื่อง (List of Subject Headings) ให้เป็น thesaurus
3. ตัดตอนรายการคำศัพท์จาก thesaurus หลาย ๆ สาขาวิชาที่ไม่ได้ให้รายละเอียดของแต่ละสาขาวิชา แล้วมาจัดทำเป็น thesaurus เฉพาะเจาะจงสาขาวิชา เช่น นำ macro-thesaurus มาตัดตอนเป็น micro-thesaurus

4. รวบรวมคำศัพท์ candidate terms จากแหล่งข้อมูลต่าง ๆ เช่น อภิธานศัพท์ พจนานุกรม สารานุกรม สิ่งพิมพ์/เอกสารอื่นในสาขาวิชานั้นและข้อคิดเห็นจากผู้เชี่ยวชาญในสาขาวิชานั้นว่า คำใดสมควรใช้เป็น descriptors และประมวลออกมาเป็น thesaurus เฉพาะสาขา

ในลักษณะการจัดทำ thesaurus ทั้ง 4 แบบนี้ พอสรุปได้ว่าแยกออกเป็นประเภทใหญ่ ๆ ได้ 2 ประเภท คือ Empirical approach และ Committee approach

Empirical approach คือ รูปแบบที่เกิดขึ้นโดยยึดหลักการทดลองและการสังเกต descriptors ได้มาจากการจัดทำตรวชนีของเอกสารกลุ่มหนึ่ง และนำเอา candidate terms มาศึกษา รวบรวมเป็น thesaurus เช่น Du Pont thesaurus ใน ค.ศ. 1959 และ descriptors ของ Mooers วิธีนี้อาจเรียกว่า Analytical approach

Committee approach คือรูปแบบที่ได้ candidate terms จากเอกสารที่เกี่ยวข้อง เช่น อภิธานศัพท์ สารานุกรม แหล่งข้อมูลอื่น ๆ หรือจากรายการคำศัพท์ที่ได้จัดทำมาแล้ว และจากคณะกรรมการผู้เชี่ยวชาญในสาขาวิชานั้น candidate terms จะถูกศึกษาทบทวนจนกว่าจะเป็นที่ยอมรับกันแล้วจึงจะนำมาใช้เป็น descriptors

ในการจัดทำ thesaurus ควรจะใช้วิธีผสมผสานกัน เพราะมีข้อดีและข้อเสียอยู่ในรูปแบบต่าง ๆ สิ่งที่สำคัญที่สุดคือ การให้ครอบคลุมสาขาวิชา และตรงกับกลุ่มผู้ใช้

## ลักษณะของ Descriptors

Descriptors ควรเป็นคำที่แสดงถึงข้อคิดที่เป็นอันหนึ่งอันเดียวกัน (single concept) อาจเป็นคำโดดหรือคำผสมก็ได้ (single or multiword descriptors) descriptors ควรแสดงถึงเนื้อหาที่เป็น important concepts ของตัวเอกสารนั้นและเป็นคำที่ใช้ในเอกสารนั้น

Multiword descriptors อาจเป็น bound terms, pre-coordinated terms ควรใช้คำผสม เมื่อ descriptors ที่เป็นคำโดดไม่สามารถจะแสดงเนื้อหาของเอกสารได้อย่างชัดเจนและเพียงพอ หรือเมื่อคำที่เป็นคำโดดใน multiword descriptors อาจนำไปผสมกับคำศัพท์อื่น แล้วก่อให้เกิด false coordination หรือเมื่อคำโดดนั้น เมื่อนำมาผสมกันแล้วก่อให้เกิดความหมายที่แตกต่างไปจากเดิม เช่น คำว่า Students, Teachers และ Student Teachers และถ้า descriptors ที่เป็นคำโดดมีความหมายกว้างจนต้องเสียเวลาในการค้นหาข้อมูล ควรจะใช้เป็นคำผสม เช่น schools-secondary schools หรือ library, automation-library automation

2. ใช้คำนามมากกว่า หลีกเลี่ยงไม่ใช่คำกริยา คำคุณศัพท์อาจใช้บ้างในกรณีที่ไม่มีความหมายที่แสดงถึง concept ได้อย่างแจ่มชัด เช่น Blind (แทน Blind people) หรือ Disadvantaged (แทน Disadvantaged people)

3. ไม่ใช่บุรพบท (preposition) “และ” หรือ “กับ” เชื่อม โดยมากมักเป็นคำโดด หรือถ้าเป็นคำผสมก็แสดงข้อคิดที่เป็นเอกเทศ เช่น library automation

4. ใช้คำนามที่เป็นเอกพจน์และพหูพจน์ตามแต่ความหมายของคำนามนั้น คำนามที่นับได้ ที่สามารถตอบคำถาม How many ได้ใช้พหูพจน์ ถ้าเป็นคำนามที่นับไม่ได้และตอบคำถาม How much ให้ใช้เอกพจน์ ส่วนใหญ่แล้วคำนามเอกพจน์จะใช้สำหรับ specific processes, properties หรือ conditions

โดยทั่วไปแล้ว descriptors จะใช้คำที่ลงท้ายด้วย “ing” สำหรับ processes และลงท้ายด้วย “ion” สำหรับ materials และ properties เช่น Concentrating-Concentration

ในกรณีคำที่ใช้ไม่ชัดเจนพออาจจะใช้วิธีการขยายความหมายของคำศัพท์ เช่น ไขว่คว้า ความหมายไว้หลัง descriptors ซึ่งถือเป็นส่วนหนึ่งของ descriptors ด้วย (parenthetical qualifiers) เช่น Mercury (Metal) และ Mercury (Planet) หรือถ้ายังไม่ชัดเจนพอ อาจจะใช้ scope note ได้โดยอธิบายรายละเอียดและขอบเขตของ descriptors นั้นใน scope note

ไม่นิยมใช้ตัวย่อ (abbreviation) เพราะตัวย่อนั้นอาจจะไม่เป็นที่รู้จักกันแพร่หลาย และไม่เป็นที่ยอมรับกัน และตัวย่อนั้นอาจจะใช้เฉพาะในเนื้อหาของเอกสารนั้นเท่านั้น และอาจต้องใช้ตัวใหญ่ capitalization และเครื่องหมายวรรคตอนต่าง ๆ ทำให้สับสนในการใช้ระบบคอมพิวเตอร์ด้วย ถ้ามีการใช้ตัวย่อ ต้องมีรายการโยง (cross references) ให้เสมอ

5. ไม่นิยมใช้เครื่องหมายวรรคตอนใด ๆ ถ้าต้องการจะให้ความหมายที่เด่นชัดขึ้นอาจใช้เป็น multiword descriptors หรือใช้ parenthetical specifiers คือเครื่องหมายวงเล็บ คำศัพท์ที่แสดงความหมายตัวท้ายจากคำศัพท์ตัวแรกเพื่อแสดงความหมายให้เด่นชัดขึ้น เช่น Institutes (Training Programs)

6. ใช้คำตรง (direct entry) มากกว่าคำกลับ (inverted entry) เช่น ใช้ Higher education ไม่ใช่ Education, Higher

### การใช้รายการโยง (Cross References)

รายการโยง (cross references) ประกอบด้วย Use UF (Used For) BT (Broader term) NT (Narrower term) RT (Related term) และบาง thesaurus อาจจะมี TT (Top term)

Use ใช้ในกรณีควบคุมคำพ้องคำเหมือนกัน synonyms near synonyms และ quasi-synonyms บางครั้งใช้เพื่อโยงจากคำศัพท์ที่มีความหมายอย่างเดียวกัน แต่ไม่ได้ใช้เป็น descriptors ให้ไปใช้คำที่เป็น descriptor เช่น

Interview

Use Interviewing

Labor

Use Labour

UF ช่วยในการกำหนดขอบเขตของ descriptors ที่ใช้ โดยบอกให้ผู้ใช้รู้ถึงคำศัพท์ที่ไม่ใช่และไม่จำเป็นต้องรู้ ซึ่งเป็น entry words แต่ไม่ใช่ descriptors

BT และ NT แสดงถึงความสัมพันธ์ตามลำดับชั้นของ descriptors (Hierarchical relationship between a genus and its species) descriptors ที่มีความสัมพันธ์กันจะแสดงความสัมพันธ์ซึ่งกันและกันออกมา (reciprocity) เช่น A เป็น BT ของ B ฉะนั้น B ต้องเป็น NT ของ A

RT เชื่อมโยง descriptors ที่เกี่ยวข้องกัน แต่ไม่ได้สัมพันธ์กันในลักษณะ genus-species relationship คำที่เป็น RT นั้นบางแห่งว่า ควรจะมีความสัมพันธ์ซึ่งกันและกัน เช่น A เป็น RT ของ B ฉะนั้น B จะเป็น RT ของ A เช่นเดียวกัน

### Hierarchy ของ Descriptors

เมื่อได้ descriptors แล้ว ต้องมากำหนดความสัมพันธ์ตามลำดับชั้นและรวมกลุ่ม descriptors ในตระกูลเดียวกัน descriptors แต่ละตัวจะต้องมี hierarchy ของตนเอง เช่น

A	ในที่นี้ A เป็น Top term
AB	AB เป็น NT ของ A และเป็น BT ของ ABC
ABC	ABC เป็น NT ของ AB และเป็น BT ของ ABCD
ABCD	ฯลฯ
ABCDE	

Thesaurus บางฉบับอาจกำหนดความสัมพันธ์เช่นนี้เพียงระดับเดียวเพื่อขจัดการลงรายการแสดงความสัมพันธ์ยาวเหยียดจนเกินความจำเป็น

### การลงรหัสประจำ Descriptors (Notation หรือ Code number)

Descriptors แต่ละตัวโดยเฉพาะ descriptors ที่เป็นรายการหลักจะต้องมีการให้รหัสประจำตัว เรียกว่า notation หรือ code number เพื่อประโยชน์ในการจัดลำดับ/จัดประเภทหมวดหมู่ของคำศัพท์ และอาจนำรหัสนั้นมาใช้สัญลักษณ์แทนคำ descriptors ต่าง ๆ ได้ notation ที่ดีควรมีลักษณะดังนี้ คือ

1. ระบบการเรียงลำดับควรเข้าใจง่าย อาจเป็นตัวอักษรทั้งหมด หรือตัวเลขทั้งหมด หรือทั้งตัวอักษรและตัวเลขประกอบกัน ควรจะจำได้ง่ายด้วย ถ้าเป็นไปได้อาจมี mnemonic features

2. ควรจะสั้นและย่อ ซึ่งขึ้นอยู่กับฐานของรหัสที่ใช้ เช่น ถ้าเป็นตัวอักษร (alphabetic notation) มีฐาน 26 ตัว A-Z จะสั้นกว่าใช้ตัวเลข (numeric notation) ซึ่งมีฐาน 10 ตัว 0-9 ถ้าใช้ทั้งตัวอักษรและตัวเลขอาจทำให้ฐานนั้นใหญ่ขึ้น และสั้นลงได้อีก

3. notation ควรจะมีการขยายเพิ่มเติมได้ทุกจุด เพื่อแสดงลำดับชั้นของ descriptors

## บรรณานุกรม

- Lancaster, F. Wilfrid. *Information Retrieval Systems : Characteristics, Testing and Evaluation*. New York : John Wiley & Sons, 1968.
- Lancaster, F. Wilfrid. *Vocabulary Control for Information Retrieval*. Washington : Information Resources Press, 1972.
- Townley, Helen M. and Gee, Ralph D. *Thesaurus-Making : Grow Your Own Word-Stock*. London : Andre Deutsch, 1972.
- UNESCO *Thesaurus : A Structured List of Descriptors for Indexing and Retrieving Literature in the Fields of Education, Science, Social Science, Culture and Communication*. Compiled by Jean Aitchison. Paris : UNESCO, 1977.
- U.S. Educational Resources Information Center. *Thesaurus of ERIC Descriptors*. 2 ed. Washington : U.S. Government Printing Office, 1969.

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย