Evaluation of Indexing

- Effectiveness
- Efficiency
- Completeness
- A good index is the result of many factors
- The history of indexing evaluation is as long as indexing

Information is Critical

- Unretrieved information is the same as nonexistent information
- Factors make impacts on information retrieval system
  - Indexing
  - File structuring
  - Coding
  - Faulty searching procedures
  - Bad computer programming
  - User interface
  - ...
- System factors and human factors affect indexing

System Factors

- Nature of index language
- Constraints of exhaustivity
- Constraints of specificity
- Level of coordination
- Overall structure

Human Factors

- Indexing consistency
- Subject expertise
- Indexing accuracy
- Indexing experience

Close-up to Human Factors

- Input and output to an information retrieval system
  - Input part
    - Consistency among multiple indexers
    - Experience of indexers
    - Subject-matter knowledge of indexers
  - Output part
    - Experience of searchers
    - Skill of searchers
    - Subject-matter knowledge of searchers
Evaluation

- A system with theoretical body could be evaluated in a quantitatively way
- Each step in indexing are similarly critical
- Too many variables cannot be dealt with in indexing evaluation together
- A controlled environment is built for indexing evaluation

Evaluation – The general problem

- What is a good index?
- Define goodness in terms of objectives
  - Does it fulfill its stated purposes?
  - Are its scope and coverage adequate?
  - Does it meet information need of users?
- Indicators
  - Accuracy
  - Consistency
  - Form
  - Internal structure

Approaches

- Evaluation of a single index
  - Needs of clientele
  - Subjects covered
  - Stated purposes
  - Cost
- Comparison of multiple indexes
  - Relative quality
  - Relative cost

Indexing Comparisons have made

- Human indexing has been inter-compared for consistency
- Human indexing has been compared with machine indexing
- Relative utility of using different parts of a document for indexing
- Statistical methods and quasi-mathematical models have been proposed to ascertain quality of indexes

The Problem

- Subjective nature of what a good index is

Milestone of Indexing Evaluation

- Cranfield I
  - Focus on indexing and searching
  - Simple model
    - Collect a set of test documents
    - Devise a search procedure
    - Submit artificial queries
    - Judge the relevance of retrieved documents
  - If the results were poor, the fault was attributed to the indexing
- Cranfield II
Controversy Results of Cranfield II

- Simple term index language give better results
- Groups of terms drop in retrieval performance while single term index language used
- Simple coordination gives better precision than more complex devices
- Simple is the best?
- Still debate

Evaluation based on User’s Need

- The user’s external expression of need may not truly express the internal need
- Users know what is needed but do not realize that they are not expressing it the way that the system requires
- Before evaluation can be carried out, some criteria of user needs and demands concerning an index must be established

Types of User Needs

- Overt information related to the item
  - Author or title
- A subject need that is specific and well-defined
- A vague and ill-defined need

Evaluation and Pertinence

- Relevance is the relationship between a document and a request
- Pertinence is the relationship between a document and an information need
- Documents are relevant to query but not pertinent to the user
  - Documents are not timely
  - Documents are in foreign languages
  - Documents are beyond the understanding of the user
  - Documents are already known

Relevance and Pertinence (continued)

- Relevance is associated with the relationship between document and index
- Pertinence is associated with the relationship between document and user
- Pertinence is concerned with the immediate usefulness to a particular user
Types of Assessors

- An information intermediary
  - Search expert
  - Know the searching strategies
  - Know how to ask
- An subject specialist
  - Subject expert
  - Know the subject matter of the request
- The requester
  - User
  - Layman

Recall and Precision

- Recall
  - The index's ability to let relevant documents through the filter
  - A ratio of the relevant documents retrieved to the total number of relevant documents potentially available
  - Measure the completeness of the output
- Precision
  - The index's ability to hold back documents not relevant to the user
  - A ratio of the relevant documents retrieved to the number of documents retrieved
  - Measure the preciseness of the output

Index

- Be easy to read
- Be detailed
- Reflect the user's viewpoint
- Have multiple entry points for an idea

Abstracts

- Represent the item's aboutness
- Exclude unimportant information
- Be error free
- Be brief and readable

Indexing

- Coverage is complete
- Consistency in term choices
- Term choices are appropriate to the nature of the users
- There is adequacy of cross-references but they are not overzealously done
- No undifferentiated subheadings
- Subheadings truly reflect the main heading
- No incorrect or missing locators
- No undifferentiated locators
- No proper names missing
- Alphabetization is consistent throughout the text
- Misspellings are all corrected
- No indexing the same topic under different index terms without proper cross-referencing
- No circular cross-references

Consider Indexes & Abstracts ONLY
Editing

- Correct alphabetizing
- Divide long multilevel headings
- Eliminate synonymous headings by consolidation under preferred term
- Recheck the editing
- Correct too many locators attached to an entry
- Correct ambiguous headings: base? Is it military, or is it a Tiffany lamp?
- Verify cross-references
- Correct spelling
- Check punctuation and capitals
- Check need to add entries
- Are locators correct?
- Are main headings relevant and/or needed?
- Is the index overdone?
- Is the index underdone?

Overall Evaluations

- Evaluate index in terms of information retrieval evaluation
- Evaluate precision and recall
- Evaluate effectiveness
- A Controlled Environment
- Test Collection
- Evaluate system, LIS, NTU Indexing & Abstracting Lecture07

Test Collections

- Test Collections
  - SMART Collections, OHSUMED, Cystic Fibrosis, LISA,...
  - BMIR-J2
  - TREC
  - NTCIR, IREX,
  - AMARYLLIS

Sample Document

- TREC: Text REtrieval Conference
- 主題: NIST及DARPA，為 TIPSTER文件計劃之子計劃之一
- 文件集
  - SGB以上
  - 基本真實文件

Sample Document
Sample Topic

```
<TOPIC>
<TITLE>金겟†</TITLE>
<DESC>ʧୗ金겟†</DESC>
<NARR>޴ᗫ˖΁不潖Œ وعد了金aget†里 записи金aget†里†_then† blackout</NARR>
<CONC>]</TOPIC>
```
NTCIR ~ 簡介

- NTCIR: NACSIS Test Collections for IR
- 主辦: NACSIS(日本國家科學資訊系統中心)
- 發展背景
  - 大型日文搜索引擎的需求
  - 跨語言檢索的研究發展需要
- 文件等
  - 來源為NACSIS Academic Conference Papers Database
  - 主要為會議論文的摘錄
  - 約330,000篇文件，其中約2/3為英文文獻
  - 部分包含part-of-speech tags

NTCIR ~ 調查主題

- 來源: 根據實際的使用者需求，再輔以修正或寫
- 每個子主題領域各100個調查主題
- 組成結構
  - <TOPIC q=nnnn>標題
  - <title>標題</title>
  - <description>資料來源及摘要概述</description>
  - <narrative>資料來源及摘要概述，包括事務上之相關性，
    類型的來源，背景資料，檢查的目的，預期的相關文件
    數量，規範的文件類型，相關判斷的標準等
  - <concepts>相關聯合的關鍵詞</concepts>

Performance Measures

- Rigid
  - S, A: relevant
  - B, C: irrelevant
- Relax
  - S, A, B: relevant
  - C: irrelevant

NTCIR ~ 相關判斷

- 判斷方法
  - 利用pooling method進行評估
  - 各含主題專家，及查詢主題的編輯者進行評估
- 判斷基準
  - S: 非常相關
  - A: 相關
  - B: 部分相關
  - C: 不相關

Standards

- ANSI/Z39.4 – 1984 Basic criteria for indexes
- ISO 999: 1996 Guidelines for the content, organization and presentation of indexes
- BS 3700: 1984 Preparing indexes to books, periodicals, and other documents
- BS 6529: 1984 Examining documents, determining their subjects and selecting indexing terms