

自由討議：

自動索引システム評価用ベンチマークテキストDBの構築

司会：石川徹也（情報大）

パネリスト：増永良文（情報大）、木本晴夫（NTT）、諏訪秀策（JICST）

テキストデータを対象に、自動索引システムの研究開発、製品の提供がなされている。しかし、その性能については、研究開発プロジェクトおよび提供システム毎に独自の評価法を基にした性能値が示されているのが現状であり、システムを客観的に評価しようとする場合に問題がある。そこで、自動索引システムの評価の為のベンチマーク用データベースを構築し、その評価法を設定をすることについて討議する。

Free Discussion:

Construction of a Benchmark Text DB for evaluating Automatic Indexing Systems

Chaired by: Tetsuya ISHIKAWA (University of Library and Information Science)

Panelists: Yoshifumi MASUNAGA (University of Library and Information Science)

Haruo KIMOTO (NTT Corporation)

Syuusaku SUWA (JICST)

The research and development of automatic indexing systems for text data has made much progress recently, and its products have been appearing on the market. Performance of the systems, however, is usually evaluated with diverse criteria adopted individually. Therefore, making objective evaluation of the systems is difficult. We will discuss about constructing a benchmark text DB and establishing an evaluation method for the automatic indexing systems.

1. 目的

テキストデータを対象に、自動索引システムの研究開発、製品の提供がなされている。しかし、その性能については、研究開発プロジェクトおよび提供システム毎に独自の評価法を基にした性能値が示されているのが現状であり、システムを客観的に評価しようとする場合に問題がある。この原因の一つは、今回の木本晴夫氏の紹介発表にあるように、アメリカではベンチマーク用データベースの提供がなされているのに、日本ではまだその整備が出来ていない点にあると考えられる。

そこで、わが国において、自動索引システムの評価の為のベンチマーク用データベースを構築し、その評価法を設定をすることは意義あるものと考えられ、それを討議する。

尚、情報検索システムの性能評価についても、依然、明確な評価法は確立していないものと考えられる。このことも関連し議論ができればと考える。

2. 話題提供

1) 自動索引システム機能の評価法の設定

i) 従来の方式例：特定テキストデータベース既付与索引語と抽出キーワード利用による検索結果の比較

ii) 評価関数：実験対象データ件数… v 、

既付与索引語利用検索結果… m 、抽出キーワード利用検索結果… n 、

有効データ（被検者判定）件数… i

・ 既付与索引語の精度

既付与索引語利用検索結果の再現率 (Recall ratio) = $m/v \times 100(\%)$ …①

既付与索引語利用検索結果の適合率 (Relevant ratio) = $i/m \times 100(\%)$ …②

既付与索引語の平均精度率 (Precision ratio) = ① / ② $\times 100(\%)$ …1)

・ 自動索引システムの精度

抽出キーワード利用検索結果の再現率 (Recall ratio) = $n/v \times 100(\%)$ …③

抽出キーワード利用検索結果の適合率 (Relevant ratio) = $i/n \times 100(\%)$ …④

自動索引システムの平均精度率 (Precision ratio) = ③ / ④ $\times 100(\%)$ …2)

・ 自動索引システムの性能評価値 (平均精度率の比較) = 1) - 2) ≥ 0

2)情報検索システム機能の評価法の設定

i)従来の方式例：特定テキストデータベースに対する検索結果の比較

ii) 評価関数：データ件数… v 、

Aシステムの検索結果… m 、Bシステムの検索結果… n 、

有効データ（被検者判定）件数… i

・ A、B両システムの性能評価値

= Aシステムの平均精度率 - Bシステムの平均精度率 ≥ 0

3)ベンチマーク用テキストデータの種類、量および質の設定

i)テキストデータの種類…日本語テキスト／英語テキスト

ii) テキストデータの量…？（統計学的最低評価用データ件数=30件）

iii)テキストデータの質…？（論文=分野？／文書？、本文／抄録？）

iv) 既付与索引語の種類および設定…主題表示索引語／ファセット索引語？

4)ベンチマーク用データのフォーマットおよび媒体の設定

i)フォーマット…SGML仕様フォーマット？

ii) 媒体…CD-ROM？

5)その他設定事項

i)既付与索引語と抽出キーワードの比較判定基準…形態一致／意味的一致？

◎「ベンチマーク用テキストデータベースの構築およびシステム性能の評価法の設定推進

方式」…オーソライズの必要性から、だれが、どこで、どの様に行うのか？

（文責 石川徹也）