

情報型クエリのためのアンカーテキスト検索モデル

Anchor Text Retrieval Model for Informational Query

学籍番号：201721710

氏名：CHEN TAIRUN

アンカーテキストはリンク先ウェブページを表す短い文字列で、リンク先の内容を簡潔に述べていて、ユーザーが他のウェブページに移動することを導くため、アンカーテキストはクエリとウェブページを繋げる重要な鍵であると思われる。従って、アンカーテキストはウェブ検索によく使われていて、アンカーテキスト検索モデルも研究されている。しかし、今までのアンカーテキスト検索モデルは案内型クエリに役立ち、情報型クエリには役立っていない問題がある。

本研究の目的は情報型クエリに役立つアンカーテキスト検索モデルを提案し、アンカーテキストの情報型クエリに対する有用性を示すことである。そして、ウェブ検索において、サイト内リンクの有用性を示すことである。本研究で提案するアンカーテキスト検索モデルをサイト間リンクだけで検索するサイト間モデル、サイト内リンクだけで検索するサイト内モデルと両方を合併した3種類に分けて評価実験を行なった。評価実験はウェブコーパス ClueWeb12 を対象に評価指標 nDCG と MAP の2つで行なった。

評価結果として、nDCG と MAP 共に、合併モデルが他の2種類のモデルより評価結果が良かった。そして、サイト内モデルの評価結果がサイト間モデルより高かった。既存のアンカーテキストモデルを実装して比較したが、提案モデルが有効であった。コンテンツモデルであるベースラインと比較しても一部の評価指標で良い結果となるなど遜色ない結果を得られた。

本研究では、情報型クエリに対する役立つアンカーテキスト検索モデルを提案した。結論として、ウェブ検索においてアンカーテキストの有用性を示したと思われる。しかしながら、今後の課題としてクエリ拡張などの性能を向上するプロセスの実装が残される。

研究指導教員：高久 雅生

副研究指導教員：于 海涛