

情報検索支援のための情報鮮度の提案

青柳 叶和

情報検索をした際に検索結果に表示される Web サイトが現在は情報の価値が低いにも関わらず、過去のアクセス数や文脈の一致具合などから検索結果の上位に表示されることで検索者が検索結果から現在も情報の価値が高い Web サイトを判別できないという問題がある。それに対して本研究では、検索時の情報の価値を数値で表した新たな尺度である「情報鮮度」を提案することで検索者に情報の価値を示し、情報検索における時間と労力の削減を目的とする。

情報鮮度は、時間的尺度、情報種別分類、期間閲覧頻度の3つの要素から構成される。最終更新日からの時間経過を測定することで、最新の情報が更新されたのか、今までの情報が更新されていないのかを時間的な観点から測定し、情報の新鮮さを定義することが可能となると考え、時間的尺度を設定する。情報は扱う分野によって経年劣化の速度が異なると考え、情報を基礎情報、応用情報、一般情報の3種類に分類した情報種別分類を設定する。検索時に多くの人が利用している情報は、情報としての価値が高いと考え、直近1ヶ月のアクセス数を基準とする期間閲覧頻度を設定する。

本研究では、情報の価値の移り変わりが激しいプログラミング分野について、情報鮮度の提案が現在の情報の価値を示すのに有効かどうかを評価する実験を行なった。評価対象の Web サイトを正しく評価する必要があるため、基礎レベルのプログラミングの知識がある学生17名を対象とした。プログラミング言語や扱う技術の異なる6つの検索テーマに対して、1テーマあたり5つの Web サイトを Google 検索から収集し、各 Web サイトの情報鮮度を算出した。算出された数値の高い順に5つの Web サイトを並べ替え5~1点の評点を付けた。一方、協力者には6テーマのうち扱ったことがあるプログラミング言語から2テーマを割り当て、5つの Web サイトを情報としての価値が高いと思う順番に並べ替えてもらった。協力者の回答データから合計点を算出し、評価が高い順に5~1点の評点を付けた。「情報鮮度」の評価による並べ替え、人の評価による並べ替え、Google の検索結果による並べ替えの3種類の並べ替えたデータを用意し、クロス集計分析と相関関係分析を行なった。その結果、情報鮮度の評価と人の評価の間には中程度の正の相関関係があった。特に、上位と下位の評価は情報鮮度の評価と人の評価の一致度は高かった。また、情報鮮度を構成する要素の1つである期間閲覧頻度と人による評価の間にも中程度の正の相関関係があった。それ以外の要素には有意な関係性はなかった。

本研究の結果から、提案する情報鮮度は既存の検索システムと比較して人にとっての情報の価値が高い情報と低い情報について、高い精度で判別することが可能であることがわかった。情報鮮度の数値を検索結果で可視化することで、Web サイトにアクセスすることなく人にとって価値の高い情報と低い情報を判別し、情報検索の時間と労力を削減することが期待される。

(指導教員 松村敦)