



Title	構造同値に基づくWWW解析に関する研究
Author(s)	山下, 長義
Citation	大阪大学, 2009, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/49703
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 ＜a href="https://www.library.osaka-u.ac.jp/thesis/#closed">https://www.library.osaka-u.ac.jp/thesis/#closed >大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

		【2】			
氏 名		やま	した	なが	よし
		山	下	長	義
博士の専攻分野の名称		博 士（情報科学）			
学 位 記 番 号		第	2 3 0 4 5	号	
学 位 授 与 年 月 日		平成 21 年 3 月 24 日			
学 位 授 与 の 要 件		学位規則第4条第1項該当			
		情報科学研究科情報数理学専攻			
学 位 論 文 名		構造同値に基づく WWW 解析に関する研究			
論 文 審 査 委 員		(主査)			
		教 授 沼尾 正行			
		(副査)			
		教 授	八木 厚志	教 授	谷田 純
				教 授	森田 浩
		准教授	栗原 聡		

論 文 内 容 の 要 旨

近年のWebの急激な発展に伴い、キーワードによる検索手法だけでは欲しい情報を効率的に収集することが困難になりつつある。これを解決するための研究として、ユーザーが従来の検索様式にて収集する情報を「手がかり」として、ユーザーの欲しい情報をWebから収集するシステムの構築を目指す研究が現在盛んに行われている。従来研究では、基本的にWebページの中身である「文字情報」を利用するものがほとんどである。これに対して、本論ではWebを特徴付ける

「リンク」に着目する。あるWebページ同士に類似性が認められる場合、そのWebページ同士がそれぞれ類似するリンク構造を持っていると考えられる。そこでWebページを分類するために、社会科学などで古くより開発されてきた「構造同値」という概念を導入した。ネットワーク内の2つの点が他のすべての点と完全に同じ関係を持つ場合、2つの点は構造同値であるという。ただし、構造同値の定義は制約が強いため、相関関係を計算する方法により構造同値の度合いを求め、それを利用してWebページを分類する手法を提案した。まず検索サイトからキーワードにより欲しい情報を含む種となるWebページを収集し、提案手法によって内容が類似するページごとに分類した。また、本提案手法を提案通りに実装すると、独自にWebにアクセスして情報を収集する必要がある処理に大幅な時間を要する。この問題に対しては検索サイトからの情報のみで分類を行う方法を提案し、処理時間を大幅に短縮させることに成功した。

文字情報を利用する方法においても、Web上の情報は新聞記事のように統制されてないため、使われている単語や表現にばらつきが多く時間とともに変化していく。しかし、自然言語処理では扱う単語や表現が限定されているという仮定に基づいて処理を行うため、Web上のデータに対しては使用に耐えないという問題がある。そこで、このような問題を解消するために自然言語処理に加えてリンク解析も利用することを考え、リンク解析によるWebページの分類結果をキーワード抽出やWeb文書の要約に適用する手法を提案した。

論文審査の結果の要旨

近年のWebの急激な発展に伴い、キーワードによる検索手法だけでは欲しい情報を効率的に収集することが困難になりつつある。これを解決するための研究として、ユーザーが検索様式にて収集する情報を「手がかり」として、ユーザーの欲しい情報をWebから収集するシステムの構築を目指す研究が現在盛んに行われている。従来研究では、基本的にWebページの中身である「文字情報」を利用するものがほとんどである。これに対して、本論文では、Webを特徴付ける「リンク」を解析することによってWebページを分類する手法を提案し、さらにその手法をキーワード抽出とWebページの要約に応用している。これらの研究成果は以下のようにまとめられている。

まず、リンク構造の類似度を基にしてWebページを分類する手法を提案している。Webページ同士に類似度が認められる場合、そのWebページ同士がそれぞれ類似するリンク構造を持っていると考えられる。そこでWebページを分類するために、社会科学などで古くより開発されてきた「構造同値」という概念を導入した。ネットワーク内の2つの点が他のすべての点と完全に同じ関係を持つ場合、2つの点は構造同値であるという。ただし、構造同値の定義は制約が強いため、相関関係を計算する方法により構造同値の度合いを求め、それを利用してWebページを分類する手法を提案し、その有効性を示している。

この手法では、独自にWebにアクセスして情報を収集する必要がある処理に大幅な時間を要するため、検索サイトからの情報のみで分類を行う方法も提案し、処理時間を大幅に短縮できることを示している。

さらに、リンク解析によるWebページの分類結果をキーワード抽出やWeb文書の要約に適用する手法を提案している。文字情報を利用する方法においても、Web上の情報は新聞記事のように統制されてないため、使われている単語や表現にばらつきが多く時間とともに変化していく。しかし、自然言語処理では扱う単語や表現が限定されているという仮定に基づいて処理を行うため、Web上のデータに対しては使用に耐えないという問題がある。このような問題を解消するために、言語処理に加えてリンク解析を利用している。提案手法とリンク解析の結果を用いない方法とを比較することで、提案手法の有効性を示している。

以上のように構造同値に基づくWebページを分類する手法を示すとともに、その手法をキーワード抽出やWeb文書の要約に適用することの有効性を示している。

以上のことから、提案手法は大きな応用性を持つものであり、情報科学の発展に寄与するものである。

よって博士（情報科学）の学位論文として価値があると認める。