

氏 名	ぜっ っ こう じ 是 津 耕 司
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	情 博 第 151 号
学位授与の日付	平 成 17 年 3 月 23 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研究科・専攻	情 報 学 研 究 科 社 会 情 報 学 専 攻
学 位 論 文 題 目	Aspect Discovery : Mining Context in World Wide Web (アスペクト発見：ワールドワイドウェブにおける文脈情報のマイニングに 関する研究)
論文調査委員	(主 査) 教 授 田 中 克 己 教 授 石 田 亨 教 授 西 田 豊 明

論 文 内 容 の 要 旨

本論文では、新しい Web マイニング技術として、ある Web コンテンツが他の Web コンテンツからどのような観点で見られているかを表す「アスペクト」という概念を提案し、Web コンテンツがリンク参照される際の文脈情報を抽出・解析し、コンテンツのアスペクトを発見する技術について述べている。本論文は以下の内容から構成される。

第1章は緒言である。Web が急速に発展してきた背景には、自分の持つ情報を世界中に誰でも自由に公開できることがある。今日では、数多くの企業や組織、あるいは個人が、自分の Web コンテンツを有している。Web の大きな特徴として、ある人が作成した Web コンテンツを他の人の Web コンテンツから自由にリンクで参照できることが挙げられる。ある Web コンテンツは他の Web コンテンツから様々な文脈で参照されている。参照の文脈は、Web コンテンツに対する様々な「観点(アスペクト)」を表しており、コンテンツの役割や用途、評判などを判断する手がかりとなると考えられる。本章では、このような考えに基づいて行った研究の動機および基本的な概念について述べている。

第2章では、関連研究、および、本研究と関連研究との関連について述べている。本研究は、Web マイニング技術の1つとして位置付けられるが、これまでのマイニング技術では、従来の情報検索技術と同様にコンテンツ属性に基づいてコンテンツを評価する方法や、Web のリンク構造を解析してコンテンツの支持率やオーソリティ度を評価することが行われてきた。これらに対し、本研究では、リンク元の周辺コンテンツによって表される参照文脈を解析し、コンテンツに対する具体的な見方(観点)を評価する点が大きく異なると主張している。また、提案手法により発見された観点(アスペクト)は、コンテンツの評判情報を評価する基本情報を提供することが可能となることについて述べている。

第3章では、Web における文脈情報の閲覧について述べている。特に、本章では、Web 上の画像を対象に、文脈情報を抽出し可視化する方法について提案している。画像の文脈は、画像が使用されている周辺の Web コンテンツによって表され、画像の中身からだけでは分からない「使われ方に依存した画像の意味情報」を表す。本章では、Web における画像の隣接周辺、構造的周辺および参照元周辺を定義し、それぞれの周辺から画像の使われ方を特徴付ける周辺範囲のコンテンツを文脈として抽出する方法を提案している。また、文脈の可視化では、類似した画像の文脈を一覧提示し、画像の類似度や文脈の類似度に基づいて一覧をソートしながら文脈の比較を行う方法について述べている。更に、文脈に基づき、類似した使われ方をしている画像を検索する方法についても述べている。

第4章は、Web コンテンツに対するアスペクト発見の具体的な方法について述べている。アスペクト発見では、まず、目的の Web ページをリンクしている個々のリンクアンカーの周辺から、あるまとまった内容を表す周辺範囲内のテキストを参照コンテキストとして抽出する。提案手法では、Web コンテンツの文書構造とリンク構造を解析し、文書要素間の包含・引用関係を示した文脈構造を発見する。抽出された参照コンテキストを類似したものの同士のグループに分類(クラスタリング)し、各クラスタ内で共通するキーワード集合から個々のアスペクト記述を生成する。提案手法により実際に発見されたアスペクトの評価を行った結果、参照の文脈を特徴付ける記述を発見できることを示す一方、意味的な文脈構造の抽出

やノイズとなる周辺コンテンツの削減などを、今後の課題として挙げている。

第5章では、アスペクト発見の応用として3つのアプリケーションについて述べている。まず、アスペクト分析では、企業や商品、あるいは個人のホームページに対するアスペクトを発見することで、Web上で皆がどのような観点でそれらに興味を持っているかを調べる方法について説明している。次に、アスペクトに基づくクロスメディア相関関係の発見では、アスペクト発見手法をWebページに含まれている画像に対しても適用することにより、ある文脈でよく参照される画像やWebページを検索する方法を提案している。最後に、アスペクトに基づくWeb検索支援では、Web検索結果をアスペクトによってクラスタリングすることにより、欲しい情報とその情報を利用する状況（文脈）の両面から検索結果を絞り込む方法について説明している。

第6章は、結言であり、本研究で行った主要な成果をまとめている。具体的には、(1)Webコンテンツの文脈情報の抽出と視覚化（閲覧）、(2)アスペクトの発見手法、(3)アスペクト発見の応用である。本研究で提案したアスペクト発見方式により、Webコンテンツの役割や用途、評判などを判断する手がかりを得られることを述べている。また、実世界の個人や企業の持つWebコンテンツのアスペクト発見は、Webを介してそれらの評判情報を調べることに有用であると主張している。さらに、本研究に関する今後の課題についてもまとめている。

論文審査の結果の要旨

本論文では、新しいWebマイニング技術として、あるWebコンテンツが他のWebコンテンツからどのような観点で見られているかを表す「アスペクト」という概念を提案し、Webコンテンツがリンク参照される際の文脈情報を抽出・解析し、コンテンツのアスペクトを発見する技術を提案している。本論文は、(1)Webコンテンツの文脈情報の抽出と視覚化（閲覧）技術、(2)アスペクトの発見手法、(3)アスペクト発見の応用から構成されている。提案しているアスペクト発見方式により、あるWebコンテンツ（あるいはWebコンテンツが代表する実世界の個人や企業）の役割や用途、評判に関する情報をその周辺コンテンツから発見することが行えるようになることを示した。本研究の主要な成果は以下の通りである。

- (1) Webコンテンツに対する参照文脈の解析という、新しいWebマイニング分野を開拓した。これまでのマイニング技術では、従来の情報検索技術と同様にコンテンツ属性に基づいてコンテンツを評価する方法や、Webのリンク構造を解析してコンテンツの支持率やオーソリティ度を評価することが行われてきた。これらに対し、本研究では、リンク元の周辺コンテンツによって表される参照文脈を解析し、コンテンツに対する具体的な見方（観点）を評価する手法を提案した。提案手法により発見された観点（アスペクト）は、コンテンツの評判情報を評価する基本情報を提供することに利用できる。
- (2) アスペクト発見により、Web上における個人や企業の見られ方を分析するという、社会情報学としての新たな可能性を示した。本研究では、実際に企業や商品、あるいは個人のホームページに対するアスペクトを発見することで、Web上で皆がどのような観点でそれらに興味を持っているかを調べることができ、それらの役割や用途、評判などを判断する手がかりを得られることを示した。Webが我々の生活に浸透していくに従い、アスペクト発見に対する要望は今後ますます高まっていくと考えられる。
- (3) アスペクトを用いて従来のWeb情報検索技術を拡張することを行った。アスペクトに基づくクロスメディア相関関係の発見では、ある文脈でよく参照される様々な種類のコンテンツの組合せ（画像やWebページなど）を発見する方法を提案した。また、アスペクトに基づくWeb検索支援では、Web検索結果をアスペクトによってクラスタリングすることにより、欲しい情報とその情報を利用する状況（文脈）の両面から検索結果を絞り込む方法を提案した。

以上、本論文は、Webマイニングに関する新規性・有用性の高い手法の提案を行っており、学術上、實際上寄与するところが少なくない。よって、本論文は、博士（情報学）の学位論文として価値あるものと認める。

また、平成17年2月3日実施した論文内容とそれに関連した試問の結果合格と認めた。