

Title	会議情報の自動編集システムの作成
Author(s)	新井, 誠亮
Citation	
Issue Date	1999-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1218
Rights	
Description	Supervisor:佐藤 理史, 情報科学研究科, 修士

修 士 論 文

会議情報の自動編集システムの作成

指導教官 佐藤理史 助教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

新井誠亮

1999年2月15日

目次

1	序論	3
1.1	背景	3
1.2	本研究の目的	4
1.3	本論文の構成	4
2	会議情報編集システムの概要	5
2.1	システム概要	5
2.2	システムの利用例	7
3	重要情報の抽出	11
3.1	fj.meetings のニュース記事の特徴	11
3.2	重要情報の抽出アルゴリズム	11
3.3	重要情報の抽出精度	17
4	適切な形式での情報提示	19
4.1	編集された情報	19
4.2	情報の内部構造の把握	20
4.2.1	内部構造把握のアルゴリズム	20
4.2.2	内部構造把握の精度	22
4.3	インターフェース	24
4.3.1	会議情報検索	24
4.3.2	表示形式の決定	26
4.3.3	表示形式に基づく情報の出力	27
5	ウェブを用いた不足情報の補完	32

第 1 章

序論

1.1 背景

現在、インターネットの利用は一般的に広く普及してきており、インターネットを通して利用できる情報は膨大になりつつある。インターネット上の情報の利点は情報の更新が素早く、リアルタイムにやりとりを行なうことすら可能で、他のメディアよりも早く最新の情報の取得が可能である。また、他のメディアはプロの手によって一度編集された情報をユーザが利用するという形態であったが、インターネットでは個人が容易に情報を公開することが可能であり、情報の多様性という面では従来のメディアにはない可能性を持っていると考えられる。

このようにインターネット上の情報は、従来のメディアにはない新しい可能性を持ったものであるが、以下のような問題点を持つ。

- 必要な情報の取得が困難
- 情報の記述形式が統一されていない

1つ目の問題点はインターネット上での情報の配置がバラバラであることである。例えばある車に関する情報がどこにあるかということについて考えてみると、Web上でメーカーのホームページで紹介されている場合、個人のホームページで紹介されている場合、特定のニュースグループでその話題が出てくる場合などさまざまである。2つ目の問題点としてはインターネット上の情報は個人が発信したものが多く、従来のメディアのようにプロの手により十分に編集されていないため、形式が統一されたものとなっていない。このため、従来のメディアより利用しづらいメディアとなっている。

現在、このような特長をもったインターネットの有効利用のために、さまざまな研究が行なわれている。例えば、情報の取得を容易にするためのサーチエンジンの技術や、自動要約の技術に関する研究が行なわれている。

そのような研究の一つに、電子ニュースのダイジェストを自動生成する研究がある。現在、電子ニュースを通じて多くの情報が流通し、多くの人々がその情報を利用している。しかし、電子ニュースを読むためのソフトウェアであるニュースリーダーは、利用者の要求に合致した、適切な機能を提供していないため、電子ニュースは読者にとって必ずしも利用しやすい情報メディアになっているとはいえない。そこで、佐藤ら [1] は電子ニュースの特定のニュースグループのダイジェストを自動生成することを試みた。このダイジェスト自動生成システムは、ニュースグループ `fj.meetings` に掲載された会議情報から、そのダイジェストを作成する。このダイジェスト作成のための中心的技術は、サマリーの自動抽出技術であり、会告記事に見られるスタイル上の特徴、言語表現パターンを利用することにより、実用に十分な精度でのサマリーの抽出を実現している。

1.2 本研究の目的

上記の研究のようにダイジェストを作成することにより、統一された表記方法を取られていない電子ニュースの記事が、一様な形式に編集された。このことにより、電子ニュースの情報はユーザにとっては非常に利用しやすく編集されたと言える。そこで本研究では、さらにニュース記事から抽出した情報を、ユーザの要求に応じて適切な形式で提供するためのシステムの作成を行なった。

電子ニュースに掲載される会告記事はダイジェスト作成に必要な全ての情報を含んでいない場合がある。このような場合、不足情報を他の情報源から得ることによって補完することが求められる。本研究では、これらの不足情報を、WWW のサーチエンジンを利用して見つけだし、それを補完する機能を実現する。

1.3 本論文の構成

本論文は本章を含む 6 章から構成される。第 2 章では会議情報システム全体の概要を説明し、第 3 章から第 5 章まではシステムの各モジュールについて述べる。第 6 章で本研究の結論について述べる。

第 2 章

会議情報編集システムの概要

2.1 システム概要

作成したシステムの構成を図 2.1 に示す。本システムは以下に示すモジュールとデータベースから構成される。

1. 重要情報抽出モジュール

fj.meetings 上の会告記事を入力として受け取り、そこから会議タイトル、場所、日時、締切といった重要情報を抽出する。

2. 内部構造把握モジュール

入力として抽出された会議の重要情報を受け取り、その情報の内部構造を把握する。

3. 会議情報データベース

会議情報を格納するデータベース。ステップ 2 の出力が格納される。

4. インターフェースモジュール

ユーザからの検索要求に対して、該当する会議情報を適切な形式に編集して提示する。

5. 会議情報補完モジュール

会議情報データベース中の会議情報において住所情報が不完全である場合に、WWW を探索し、不足情報を補完する。

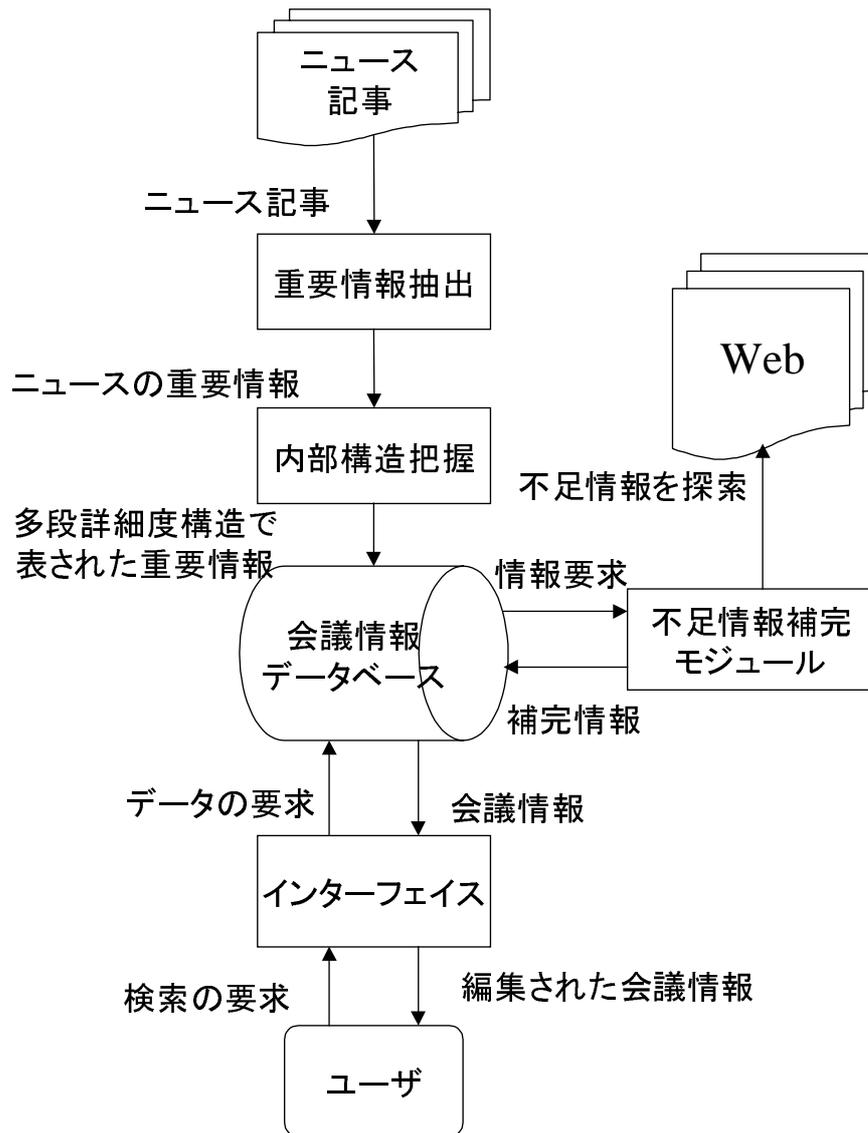


図 2.1: システム概要

2.2 システムの利用例

ユーザの検索要求入力画面を図 2.2 に示す。この画面上でユーザは探したい会議のタイトル名、場所、日付を指定することができる。システムはユーザの検索入力に対して、該当する会議情報を適切な形式に編集して提示する。その出力例 (ダイジェスト形式) を図 2.3 に示す。この図において、会議のタイトルをクリックすると、図 2.4 に示されるように、サマリー情報がページ上部に表示され、下部に元記事を表示する。サマリー情報の場所情報の部分は補完が行なわれた場合、補完元ページへリンクさせ、ユーザが補完された場所情報が正しいかどうか確認することが可能である。

