

宿泊サイトに関する分散情報の統合整理を行うシステムの提案

Proposal of the system for consolidation of distributed information about the accommodation site

神田 明範*1 大和田 勇人*2
Kanda Akinori Ohwada Hayato

*1 東京理科大学 大学院 理工学研究科 経営工学専攻
Department of Industrial Administration, Graduate School of Science and Technology, Tokyo University of Science

*2 同 理工学部 経営工学科
Department of Industrial Administration, Faculty Of Science and Technology

本研究は web 上に分散した情報について、統合と整理を行った上でユーザーに提供するシステムの提案を行う。今回は宿泊サイトのプラン情報に着目した。プラン情報は宿泊サイトごとに料金やサービスが異なっている。それらプラン情報を統合しユーザーに提供するためのシステムの提案を行う。またユーザーのタイプ別にプランの推薦を試みる。

1. はじめに

近年、急速に拡大したインターネットの情報をまとめるためのサイトが登場している。それは、一部のジャンルに対してユーザーの求める情報だけを提示できるような Web サイトである。それらの Web サイトは多数のサイトの情報を収集し、一括して表示することが可能である。本研究では、このようなサイトを情報統合サイトと呼ぶ。

本研究はそのような情報統合サイトを、宿泊サイトに特化して作成する。宿泊サイトとは、じゃらん*1に代表されるような Web サイトのことである。ユーザーは、宿泊予約サイトに登録された宿のデータから、希望の宿とその宿についての宿泊プランを選んで予約を取ることが出来る。宿泊プランは宿泊予約サイトによって異なっており、それらと比較するためにはいくつかの宿泊サイトを閲覧しなければならない。それらを統合して一括表示することにより、ユーザーは単一のページで宿泊プランの比較が可能となり、それによって検索の手間を減らすことが出来る。

そこで、本研究では、宿泊サイトのプラン情報を統合し、ユーザーのタイプ別にプラン情報を提示することができるシステムについて提案を行う。

2. 関連研究

情報統合の研究は今日まで様々な分野で研究されており、特に論文データや特定の Web サイトの情報を対象とした情報統合に関する研究が盛んにおこなわれている。例えば、Krallinger ら [1] は PLAN2L というシステムを発表している。PLAN2L は生物学的な論文を Web 上から検索する際に用いることができる。このシステムを用いることによって検索された論文が特定の分野に対してどれだけ関連性があるのかをスコアで定量的に知ることが可能である。また、于ら [2] は Web 検索エンジンを利用することによりラッパーで抽出したデータの相関を発見し、複数のデータベースを統合するシステムの提案をしている。McKeown ら [3] は複数のニュースサイトの情報を一つのページに統合して表示するシステムを発表している。それは、

連絡先: 神田 明範, 東京理科大学大学院 理工学研究科
経営工学専攻, 千葉県野田市山崎 2641, 04(7124)1501,
j7412607@ed.tus.ac.jp

*1 <http://www.jalan.net/>



図 1: プラン情報の例 (じゃらんの場合)

ニュースをただ収集するだけではなく、収集したページを複数のカテゴリに分類し、ニュースの内容を要約するシステムである。

観光業界と Web の関係として斎藤 [4] の研究にもあるように、利用者が宿泊施設の評判を Web を通じて入手したり、宿泊施設の運営者がブログ等を通じて利用を促すといったように観光業界と Web が切り離すことのできない関係となっているのが現状である。

3. 提案システム

3.1 宿泊プランとは

宿泊プランとは、その宿泊施設に宿泊する際に「どのような部屋に泊まるか」、「食事の有無」、「チェックイン、チェックアウトの時間」などが設定された利用者がその宿泊施設でどのようなサービスを受けることができるのかを表したものである。例えば、図 1 はじゃらんにおけるプラン情報の一例である。本研究ではプラン名を大プランと名づけ、ベッドタイプや食事の有無など、詳細な情報と料金情報を小プランと名付けた。

本研究ではプラン情報を「一人」、「友だち」、「カップル」、「家族」の 4 タイプに分類し、利用者に分類した結果を表示することを目標とする。

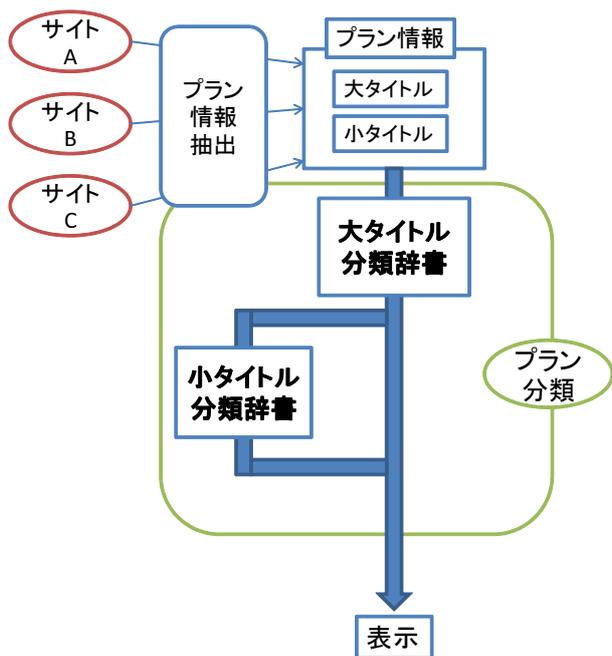


図 2: システム概要

表 1: 分類辞書内のキーワード

タイプ	キーワード例
一人	「1人」, 「一人旅」
友だち	「仲間」, 「女子会」
カップル	「彼氏」, 「彼女」
家族	「家族」, 「ファミリー」

3.2 システム概要

提案システムは図 2 のようになる。まず各宿泊サイトから宿泊施設に関するプラン情報を抽出し、それらプラン情報から大タイトルと小タイトルを判別する。次に大タイトルを分類辞書に通し、プラン情報を分類することを試みる。大タイトル分類辞書には表 1 のようなキーワードが大タイトルの中に含まれていた場合、分類を行っていく。大タイトルでの分類が行えなかった時は、小タイトルからの分類をおこなう。小タイトルの中でも特にベッドタイプに注目し、例えばシングルタイプの場合なら「一人」、ダブルベッドタイプなら「カップル」というように分類を行っていく。

4. 今後の課題

現状このような分類を行っているが大タイトル、小タイトルで分類が行えなかった場合はそのプランが未分類となってしまう問題がある。そのような場合、方法の一つとしてプランの詳細情報から手掛かりとなる情報を見つけたことである。例えば、図 3 のように詳細情報から部屋の広さや「眺めがよい」、「レジャー施設のチケット付き」などといった情報を知ることができる。それらの情報を用いてより詳細な辞書を生成することにより、正確な分類が可能となる。

また、その際に機械学習を用いれば、より正確な分類辞書の

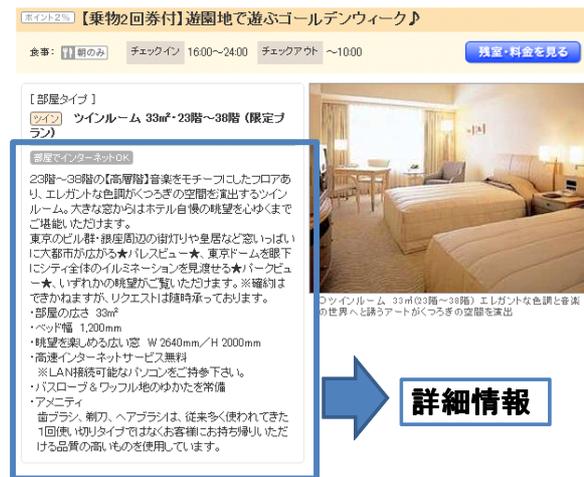


図 3: プランの詳細情報

生成が可能ではないかと考えられる。

5. まとめ

本研究は宿泊サイトを対象としたプラン情報の統合分類システムを提案した。その際に分類辞書を用いることで宿泊プランを自動的にタイプ別に分類した。今後の課題としてはより詳細な情報を用いることにより正確な分類が期待できると考えられる。

参考文献

- [1] Martin Krallinger, Carlos Rodriguez-Penagos, Ashish Tendulkar and Alfonso Valencia "PLAN2L: a web tool for integrated text mining and literature-based bioentity relation extraction", Nucleic Acids Research, 2009, Vol. 37, Web Server issue p.p.160-165 (2009)
- [2] 于 美麗, 天笠 俊之, 北川 博之「レコード間の相関を用いた Web 情報統合方式の評価」, 情報処理学会第 73 回全国大会 p.p.569-570 (2011)
- [3] McKeown, K.R., Barzilay, R., Evans, D., Hatzivassiloglou, V. Klavans, J.L., Nenkova, A., Sable, C., Schiffman, B. and Sigelman, S. "Tracking and summarizing news on a daily basis with Columbia's Newsblaster", Proceedings of HLT'02 (2002)
- [4] 斎藤 一「Web における観光情報提供と分析」, 人工知能学会誌 26 巻 3 号 p.p.234-239 (2011)