

ドキュメント群を利用した設計プロセス支援

Using document set to support design process

唐澤 悠紀*¹ 田中 克明*² 赤石 美奈*² 堀 浩一*²
Yuki KARASAWA Katsuaki TANAKA Mina AKAISHI Koichi HORI

*¹東京大学大学院工学系研究科航空宇宙工学専攻

Department of Aeronautics and Astronautics, School of Engineering, The University of Tokyo

*²東京大学先端科学技術研究センター

Research Center for Advanced Science and Technology, The University of Tokyo

In this paper, we propose a user interface system which supports users to find various kinds of relations among a lot of information. We discuss the function of an information sharing system from the viewpoint of relations among information. We try to construct "Personal Dictionaries" and "Master Dictionary". "Personal Dictionaries" are dictionaries which contains related words of each user. "Master Dictionary" is a union of all "Personal Dictionaries". They help users to get related words among information.

1. 研究の背景

航空宇宙をはじめとする大規模システムは年々発達してきているが、その失敗事例は途絶えることがない。例えばスペースシャトル・コロンビア号の空中分解事故、我が国でも日航機の前輪タイヤ脱落、全日空機の高度誤認など航空機のトラブルが相次ぎ、H-IIA ロケット 6 号機の打ち上げ失敗もいまだ記憶に新しい。このような失敗は、設計時の想定を超えた負荷や予期せぬ事態が発生したことに起因するという側面もある一方で、設計開発の過程で得られる情報や既存の知識や経験を正しく利用することで防げたものも少なくない。このような状況に対応するため、多くの情報管理、獲得、構造化技術が研究され [Baeza-Yates 99]、情報の共有が試みられている。

そこで我々は情報を共有し、共有した情報の関係を提示するシステムを提案する。本研究では東京大学大学院工学系研究科航空宇宙工学専攻中須賀研究室において行われている、超小型衛星 CubeSat "XI-IV" *¹ の設計過程において作成された議事録を対象とし、個々の情報の関係を提示することを目的として、支援システムを提案し、現在開発を行っている。本稿では提案するシステムの設計について述べる。

2. 情報間の関係提示

情報共有技術に求められる事の一つとして、「共有した情報間の関係の提示」が挙げられる。大きなシステムを扱う場合、扱われる情報は膨大な量となり、システム全体を把握することは難しくなる。そこでシステムは細分化され、部分ごとに専門のグループを作り対応することになる。しかし、細分化されたグループが個々に開発や設計、仕様の変更を行った場合、その影響はグループの専門領域に留まらずシステム全体へ影響することになる。個々の影響は小さくとも、複数の要因が重なることで事故へと繋がるのが多々ある。

そのような事故原因のひとつとしては、情報の共有化が不十分だったこと挙げられる。しかしながら、情報を共有して

も個々の情報間の関係を把握できていなければ、膨大な量の情報に埋もれてしまい、有効に利用することができない。過去に白川が、超小型衛星プロジェクトを手がける研究室のメンバーに対して行ったアンケート [白川 02] でも「蓄積された情報の中から目的の情報を探すときに、自分で検索範囲を絞ってしまい、適当な情報を見つけない」という問題が指摘されている。

このように情報は共有するだけではなく、グループ内の関心に関係した情報を適切に提示する必要がある。

3. 提案システム

情報間の関係の一つとして「単語による繋がり」が挙げられる。例えば、多くのドキュメントの中で同じ単語を用いているものは、少なからず関係があるということは容易に想像が付く。また違う単語でも、同じような意味のものや関連する単語を用いたドキュメント同士も関係があるだろう。だが情報共有の観点から見た場合、そのような異なる単語による繋がり提示は容易ではない。

共有した情報を利用する場合、一般的には目的の情報に関連する単語で検索をかける。目的の情報を得られなかった場合は、再度別の関連する単語で検索するだろう。しかし、関連する単語を個人で考えるには限界があり、また個人ごとの表現の違いによって関連する情報が得られないこともある。

本研究では個人ごとの表現の違いに着目し、自分では思いつかなかった「他人の表現」を提示することによって、情報間の関係の把握を支援するシステムを提案する。本稿では他人の表現を抽出するために共起依存度 [赤石 06] を用いる。

3.1 システムの構成

システムは図 1 に示す 4 つのパートで構成される。システムはまず、ドキュメント群から各個人の使用する単語関係を数値化して個人辞書を作成する。つぎに作成された個人辞書全ての和集合をとりマスター辞書を作成する。そして提示システムがドキュメントの内容と、マスター辞書内の関連語を提示する。これによって、表現の違いによる見落とししていた情報も、同じ関連語で繋がっている可能性が高いので見落とす可能性が減少すると考えられる。また他人の関連語が提示されることにより、ユーザ自身では思いつくことができなかった情報の関係が得られ、プロジェクト等の失敗回避に繋がると考えられる。

連絡先: 唐澤 悠紀 (KARASAWA Yuki), 東京大学大学院工学系研究科, 〒153-8904 目黒区駒場 4-6-1 東京大学先端科学技術研究センター 知能工学研究室, karasawa@ai.rcast.u-tokyo.ac.jp

*¹ <http://www.space.t.u-tokyo.ac.jp/cubesat/>

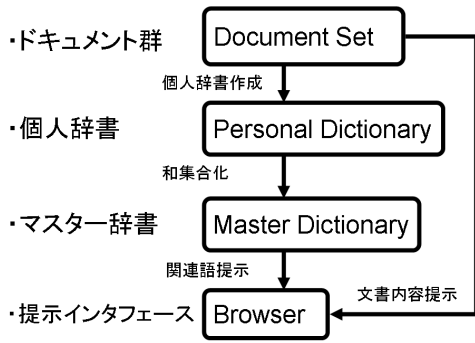


図 1: システム構成

3.2 ドキュメント群

ドキュメント群は個人ごとの関連語を抽出する為に、筆者の情報を付加して管理する。ドキュメント群の概略を図 2 に示す。筆記者は通常個人として捉えて情報を付加するが、筆記者をグループや時間として捉える事も可能である。その場合、ドキュメント群はグループや時間の情報を付加して管理することとなり、抽出されるのはグループごとの関連語、時間ごとの関連語となる。

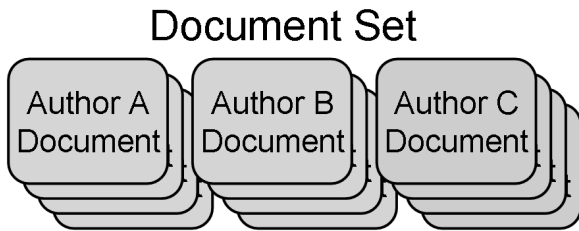


図 2: ドキュメント群

3.3 個人辞書

ドキュメント群から個人ごとの関連語辞書である「個人辞書」を作成する。個人辞書の概略を図 3 に示す。まずドキュメント群の中から筆記者による抽出を実行し、茶筌 [松本 00] を用いて形態素解析を行って単語を取得する。このとき茶筌の形態素コストによってフィルタリングを行い、助詞や助動詞などの特徴に乏しい単語を排除する。取得した単語全てに対して式 (1) から式 (3) を用いて共起依存度 $td(t, t')$ を計算する。

単語 t を含む文の数

$$sentences(t) \quad (1)$$

単語 t と単語 t' を含む文の数

$$sentences(t, t') \quad (2)$$

単語 t の単語 t' に対する共起依存度

$$td(t, t') = \frac{sentences(t, t')}{sentences(t)} \quad (3)$$

各単語ごとに双方向の共起依存度 $td(t, t')$, $td(t', t)$ を計算し、その値に基づき以下のように関連語定義を行う。

1. 双方向の共起依存度が高い場合

単語 t と t' は文書群の中で非常に密接に関係していると考えられる。そこで $td(t, t') > \sigma_1$ かつ $td(t', t) > \sigma_1$ のとき、単語 t と t' は互いをひとまとめの”Set”関係の関連語と定義する。

2. 一方方向の共起依存度をが高い場合

単語 t と t' は文書群の中で一方方向の依存関係にあると考えられる。そこで”Set”関係語となる条件を満たさず、 $td(t, t') > \sigma_1$ のとき単語 t を t' の”Child”関係の関連語と定義する。また、 $td(t', t) > \sigma_1$ のとき t を t' の”Parent”関係の関連語と定義する。

3. 中程度の共起依存度の場合

単語 t と t' は文書群の中で中程度の依存関係にあると考えられる。そこで $\sigma_2 < td(t, t') < \sigma_1$ または $\sigma_2 < td(t', t) < \sigma_1$ のときに、単語 t と t' は互いを”Friend”関係の関連語と定義する。

以上の関連語をまとめて個人辞書とする。筆記者ごとに異なるドキュメントを用いるので、筆記者ごとに異なる個人辞書が生成される。計算で得られる共起依存度による t と t' の関係を表 1 に示す。共起依存度が高い語に関してはユーザ自身が容易に関係を想起できるので、特に見落とししやすい”Friend”関係の関連語をユーザに提示することが重要であると考えられる。

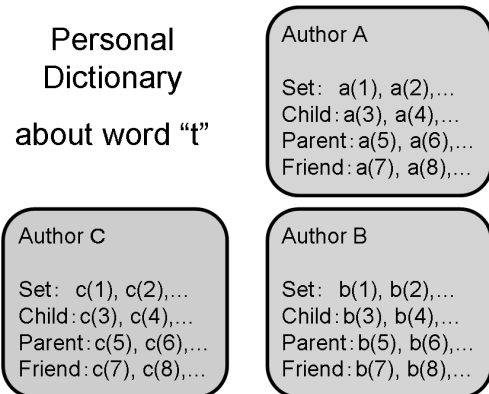


図 3: 個人辞書

表 1: t と t' の関係

		$td(t', t)$		
		Low	Middle	High
$td(t, t')$	Low			Child $t \supset t'$
	Middle	Friend		
	High	Parent $t \subset t'$		Set $t \equiv t'$

Low : $td < \sigma_2$

Middle : $\sigma_2 < td < \sigma_1$

Low : $\sigma_2 < td$

3.4 マスター辞書

全ての個人辞書の和集合をとってマスター辞書を作成する。マスター辞書の概略を図4に示す。積ではなく和をとることによって筆者全ての使用する関連語を網羅することができる。

Master Dictionary

Set:	a(1), a(2), b(1), b(2), c(1), c(2)...
Child:	a(3), a(4), b(3), b(4), c(3), c(4)...
Parent:	a(5), a(6), b(5), b(6), c(5), c(6)...
Friend:	a(7), a(8), b(7), b(8), c(7), c(8)...

図4: マスター辞書

3.5 提示インタフェース

提示インタフェースの概要を図5に示す。提示インタフェースはドキュメントの内容に対して、使用されている単語とマスター辞書から得られた関連語を提示する。また、関連語に対してその関連語や関連語を含むドキュメントのリストを提示することにより、ドキュメント同士の関係が得られると思われる。しかしながら関連語の提示方法については思案中であり、ドキュメント同士の関係を得る為には関連語を効果的に提示する方法が必要である。

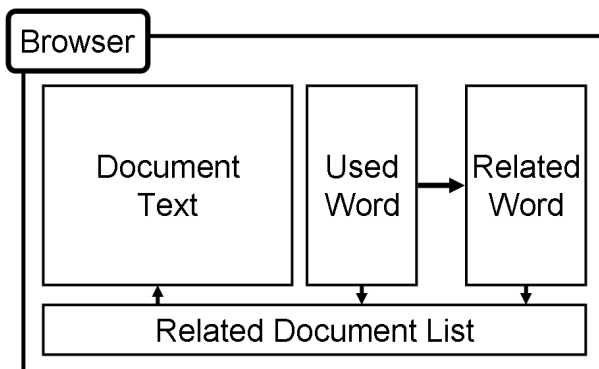


図5: 提示インタフェース

4. 今後の予定

現在提案したシステムの実装を行っている。実装後、先述の CubeSat“ XI-IV ”の設計議事録を用いて実験を行い、システムによってどのような関係を示すことができるか確認する。またドキュメントに対する共起依存度 $td(t, t')$ の適切な閾値 σ_1, σ_2 を調べる。これらを通して、ドキュメント群へ筆者以外のタグ付けを行った場合の利用方法、果的な提示インタフェースの仕組み、提示する単語量の調節方法、ドキュメント間の関係を視覚的に表示させる方法などの問題を考察していく。

5. まとめ

本稿では情報間の関係に着目して、情報共有システムを提案した。システムは、情報間の関係を個人で捉える限界や個々の表現の違いによる問題に対応する為に、「他人の表現」を集約して個人辞書として、その和であるマスター辞書が関連語を提示するものを提案した。辞書作成には共起依存度を用いることとした。今後はシステムの実装、関連語の提示方法の考案、ドキュメント間の関係を視覚的に表示させる方法の模索、実験を行う。

参考文献

- [Baeza-Yates 99] Baeza-Yates, R. and Ribeiro-Neto, B.: Modern Information Retrieval, Addison Wesley (1999).
- [白川 02] 白川敬寛, "設計情報の統合的利用による人工衛星の不具合解析支援", 東京大学修士論文 (2002).
- [赤石 06] 赤石美奈, "文書群に対する物語構造の動的分解・再構成フレームワーク", 人工知能学会論文誌 21 巻 5 号 A, pp428-438 (2006).
- [松本 00] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸, 日本語形態素解析システム『茶筌』 version 2.3.3 使用説明書, NAIST Technical Report (2000).