JAIST Repository
https://dspace.jaist.ac.jp/

Title	蓄積情報からの変化の抽出と再構成　小型衛星設計と個人行動履歴を例に
Author(s)	田中，克明；堀，浩一
Citation	第六回知識創造支援システムシンポジウム報告書: 105-112
Issue Date	2009-03-30
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/7979
Rights	本著作物の著作権は著者に帰属します。
Description	第六回知識創造支援システムシンポジウム，主催：日本創造学会，北陸先端科学技術大学院大学，共催：石川県産業創出支援機構文部科学省知的クラスター創成事業金沢地域「アウェアホームのためのアウェア技術の開発研究」，開催：平成21年2月26日～28日，報告書発行：平成21年3月30日

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

蓄積情報からの変化の抽出と再構成

小型衛星設計と個人行動履歴を例に

Extracting and Reorganizating Transitions

in Design Records of a Small Satellite, and Personal History

田中 克明
Katsuaki TANAKA
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
kcsss2009@katsuaki-tanaka.net

堀 浩一
Koichi HORI
東京大学大学院工学系研究科
School of Engineering, The University of Tokyo
hori@computer.org

Summary

When creating something, the creator is applying his or her own knowledge to an object. The object is changed by the applied knowledge. Although it is difficult to illustrate what knowledge has been applied, observation of transitions of an object in chronological order may give clues about the applied knowledge. We focused on the meeting minutes of a design task that was performed by a group of people, and tried to acquire transitions of topics in documents. An acquired transition may seem correct, but it can be difficult for someone to understand because it does not belong to anyone. Therefore we also proposed reorganization of the transition to match a reader's viewpoint to make it closer to the reader's knowledge. We applied our method to records of a design project for a super-small satellite, and personal histories in the *My Life Assist Service* that was a service for mobile phones provided by NTT DOCOMO, Inc. as a field experiment from Dec. 2008 to Jan. 2009. The acquired transition was confirmed to be approximately correct by members of each project.

1. は じ め に

近年の計算機とそのネットワークの普及と拡大に伴い，そこには膨大な情報が蓄積されている．この情報から，人間が新たな活動を行うために有用な知識を手に入れ，その知識を活用することを目的とし，計算機と人間が協調して情報に対処する様々な手法が試みられている．

多くの手法は，対処を行う時点までに集積されたすべての情報，あるいは，ある時点で得られた情報を対象として処理を行う．例えるならば，年輪を積み重ねた木が持つ成長過程に注意を払わず，ある時点の木の写真のみを見ているかのようである．

木がどのようにしてそれぞれの時点の姿となったかは，木の成長過程を撮影した写真を時間経過に沿って並べて見ると分かるだろう．つまり，ある時点での情報とは，ただそれだけが存在するものではなく，そこに到るまでの変化の積み重ねにより成り立っている．

筆者らは，この積み重なりを解きほぐし，人間が新たな積み重ねを作るための支援を行うことを目指して研究を行っている．すなわち本研究では，長期間にわたる人間の活動の結果として蓄積された情報に対し，そこに内包されている時間経過に伴う変遷へ着目し，この情報構造遷移の抽出と再構成を行うことにより，これらの活動を支えた知識を，活動に参加，あるいは周辺に位置する人間が，それぞれにとって有用な形として再発見することの支援を目的とする．

情報を扱う知的な処理のほとんどは，情報を収集し，これを分析し，新たな情報を構築し，これを適用して何かを生み出す，というサイクルである．計算機があますことなく自律的にこのサイクルを成立させることはまだ難しく，また，人間が計算機を用いずにこのサイクルを成立させることも難しくなった．

計算機を用いて情報と知識を扱う多くの研究は，知識を得る，知識をためる，知識を使うという３つの段階に知識を取りまく状況を分割し，手法を中心に各段階ごとの目的を設定するか，あるいは特定の状況の下に，各段階を横断する目的を設定している．

例えば，知識工学，データマイニングや知識発見は知識を得ることを目的とした研究であり，知識ベース，オントロジーは知識を貯めることを主目的とした研究と言える．知識を使う研究としては，エキスパートシステムがあげられるだろう．一方，知的 CAD などの設計支援システム，オントロジーの開発と利用を行うシステム，ナレッジマネジメントシステムなどは，３つの段階を含みある状況下でその応用を探る研究であると言えよう．

これらの研究に共通することは，人間と計算機が協調して情報を取り扱うにあたり，計算機が処理している対

象を人間側が形を持ったものとして確認したいという欲求をもち，これを「知識」という「確認できるもの」とするという思想を根底に横たえていることであろう．すなわち,「知識」を人間と計算機の分界点とし，それをもって相互協調を成し遂げようとする試みであるといえる．

他方，人間と計算機の協調の分界点を知識に置かず，計算機，あるいは人間が主導して情報を扱う研究もある．たとえば機械学習では，正解データを元に計算機がデータの分類方式を自律調整し情報を処理する．処理を主導するのは計算機であり，人間と計算機の分界点はデータにある．一方，創造活動支援においては人間が自身の思考として情報の処理を主導し，計算機は人間が表現したデータをそれぞれの方式で処理して人間に提示することにより，人間の支援を行う．これらの試みでは，協調分界点としての陽たる知識のあらわれはない．

計算機が知識を「使う」領域は，知識を分界点として用いる研究において，エキスパートシステム以後，それほど盛んに行われていない．その理由の一つとして，分界点となるべき知識自身が「どこに適応されるべき」かという状況の設定を内包してしまい，それを適応する局面をあらかじめ外部から十分に絞り込んみ設計を行ったシステムでないと，計算機が知識を利用する状況を絞り込めないことがあげられるだろう．知識の利用局面を限定することによりエキスパートシステムがある程度の成功を収めたこと，知識の体系化すなわち利用されるべき状況との対応付けを可能とするオントロジーの研究が近年盛んなことも，同様の理由によると考えられる．すなわち，知識を形を持ち確認できるものとしてとらえ，そのまま切り出し，別の状況に適用可能なものとして扱うことが，これらの研究の特長であり課題でもあろう．

これに対し筆者らは，知識は直接「確認できるもの」ではなく，適用した結果のみを外界からとらえることができるものであると考える．すなわち，知識自体をとらえることはできないが，知識が何らかの対象に適用されることにより，対象とその周辺の時間経過に伴う変化として，その存在を認識することができるのではないかと考えた．

そこで，時間経過に沿って蓄積された文書群から，文書がどのような内容を含んでいるのか（図図 1(a)）ではなく，構造がどのような経緯を経て作られたのか，その変化を抽出し（図図 1(b)），変化の再構成を行うことにより，人間が対象文書群から自身が必要とする知識を得る支援を試みてきた [Tanaka 04, Tanaka 08].

2. 関 連 研 究

時間経過に沿って情報を扱う研究は，例えば専用の支援システムを情報処理過程で使用することにより操作を記録し知識獲得を目指す研究 [妻屋 02][武内 07] などがあげられる．また，専用のシステムの記録を用いるのではなく，通常の文書から主題とその時間変化を抽出する研究として，例えばニュース記事を対象とした topic detection and tracking (TDT) [Allan 02]，同じく時系列にそって新聞記事などを対象とし話題の大まかな遷移を可視化する ThemeRiver[Havre 00]，統計的な情報の通時的変化の抽出と可視化を主に行う情報編纂 [松下 09] なども行われている．

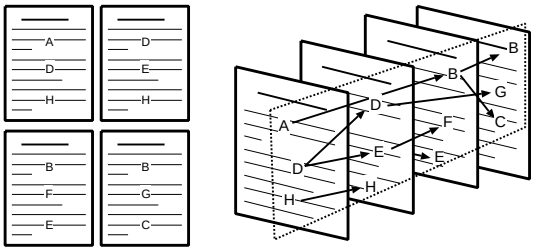

図 1 主題抽出 (a) と主題変化の抽出 (b)

3. 例 1：小型衛星設計議事録

まず，小型の人工衛星設計・運用プロジェクトの記録を対象とし変化の抽出と再構成を試みた例について述べる．

3·1 CubeSat プロジェクト

CubeSat プロジェクト*1 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（ISSL）による学生を主体しとした小型人工衛星の設計，運用プロジェクトである．これまでに XI-IV (2003)，XI-V (2005) の打ち上げに成功し，現在もあらたな衛星の設計に取り組んでいる．しかしながら，学生がプロジェクトの主メンバーであるため，必然的にメンバーが「卒業」してしまい，過去の設計で経験をつんだ設計者が現在はわずか数名となり，あらたなメンバーへの経験の伝達が課題となっている．

一般的な衛星設計に必要な情報を集めたハンドブック [Larson 99] は存在するが，ISSL が実際に経験してきた事象について,「何を作ったか」ではなく「何かを作る上でどのようなことを行ったか」を示す，いわゆるプロセス知 [吉川 97] を伝達することが現場で作業を行う上では必要であり，この目的にあう独自のハンドブックなども検討されている．

3·2 対 象 文 書 群

CubeSat プロジェクトにて，2000 年 1 月 5 日から XI-IV 打ち上げ前の 2002 年 12 月 12 日まで間に作成された，議事録，マニュアル，実験記録など，約 600 の文書を対象とし，そこに含まれる主題とその変化の抽出を試みた．各文書は，作成日付，タイトル，記録者などのヘッダ部，

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

および議事内容からなる．日付，タイトル，記録者以外の部分には統一された書式はなく，記録者により，まちまちの形式で記述されている．

3·3 変化の抽出

以下の手順により，文書群からそこに記述されている主題変化の抽出を行う [Tanaka 08].

(1) 文書作成時刻による文書集合定義
(2) 文書の断片化
(3) 文書断片のクラスタリングによる主題抽出
(4) 主題（クラスタ）の忘却処理
(5) 主題間の関連度計算
(6) 主題遷移のグラフ化

1 により時間方向の視点を文書群に与え，2, 3, 4 により主題の抽出を行い，5, 6 により抽出した主題の時間経過にともなう遷移の構造化を行う．

§1 文書作成時刻による文書集合定義

文書群 D に対して，もっとも古い文書の作成時刻と最新の文書の作成時刻の間を N 等分し，文書集合の時間間隔 S を定義する．$E(D)$ を D 中の最初の文書の作成時刻，$L(D)$ を D 中の最新の文書の作成時刻とし，文書集合の分割時間間隔 S を，以下のように定義する．

$$S \equiv \frac{L(D) - E(D)}{N}$$

S にもとづき，N 個の文書集合 $D_1, D_2, \cdots, D_N$ を以下のように定義する．$c(\mathbf{d})$ を文書 $\mathbf{d}$ の作成時刻とする．

$$D_i \equiv \{\mathbf{d} \mid c(\mathbf{d}) \leq E(D) + i \cdot S\}$$

この結果，各文書集合は，$D_1 \subseteq D_2 \subseteq \cdots \subseteq D_N = D$ となる．ここでは，処理結果を表示した際に人間が把握できる範囲を考慮して，$N = 50$ とした．

§2 文書の断片化

文書は，1 つ以上の話題を含んでいることがほとんどである．例えば，ある CubeSat の設計議事録は，全体の進捗報告，各設計担当ごとの課題，ある特定の話題についての詳細な議論を含んでおり，それぞれを別の主題へ属するものとして取り扱う必要がある．すなわち，文書を主題ごとに分割する必要がある．

本研究では，文書をある主題を記述していると思われる長さより短い長さに断片化し，クラスタリングにより似た断片をまとめ，まとめられた断片集合をひとつの主題として扱う．

文章を計算機により断片化するには，文単位で切る，あるいは形式に関係なく一定長で切るなどの単純な方法，TextTiling[Hearst 97] のように文章表現を解析する方法などがある．今回は，処理対象の文書が定められた記法を持たずまた多数の人間により記述されたものであり，表記が様々である上に句点と見なせる記号間の長さも様々であること，また，断片化されたテキストは次のクラスタリング処理により分類を行うため，詳細な解析を伴う処理までは不必要と考えられることから，一定長（W bytes）ごとの切り出しによる断片化を行った．

分割時に分割点が単語上にある場合は，その単語を断片内部に含むよう，W を必要なだけ拡張し，ひとつの単語が分割されてしまわないようにした．また，主題が複数の断片化に分割された場合には，断片中の隣接断片との重なり部分と非重なり部分が類似した内容を持つことにより，クラスタリングにより同一のクラスタに分類される可能性が高くなるように，連続する断片間には $\frac{W}{3}$ の重なりをもうけた．

§3 文書断片のクラスタリング

各文書断片に対して，茶筌*2 によって形態素解析を行い，名詞と品詞分類された単語をとりだし，単語とその出現回数からなる文書ベクトルを作成した．文書断片から作成した文書ベクトルに基づいて，文書集合ごとに，K 個のクラスタへのクラスタリングを行った．クラスタリングの結果得られた K 個のクラスタが，各文書集合 D_i における K 個の主題を表す．

クラスタリングには汎用連想計算エンジン GETA*3を用いた．GETA は，単純リンク法，完全リンク法，群平均法，Ward 法，および確率的なクラスタリング手法（HBC 法）によるクラスタリングの機能を持つ．クラスタリング時に，各単語には自動的に tfidf による重み付けが行われ，クラスタリング対象の要素間の距離にはコサイン距離が用いられる．

§4 主題（クラスタ）の忘却処理

何度も触れられる主題は設計者にとって重要なものであり，言及がない主題は，次第に忘れられつつある重要ではない主題である．

そこで，D_{i+1} に属する文書断片のクラスタリングの終了後，D_i には存在せず D_{i+1} には存在する文書断片（すなわち新たな断片）を含まないクラスタ $C_{i+1,k}$ に属する文書断片に対し，D_{i+2} 以降のクラスタリングに用いる文書ベクトル作成の際，単語の重みを R 倍 $(R < 1)$ することにより，古い主題の重みをじょじょに減らす，主題の忘却処理を行った．

§5 主題（クラスタ）間の関連度計算

隣接する文書集合間 D_i, D_{i+1} に属する主題（クラスタ）$C_{i,k}, C_{i+1,m}$ 間の関連度として，$sim(C_{i,k}, C_{i+1,m})$ を以下のように定義した．

$$sim(C_{i,k}, C_{i+1,m}) = \frac{|C_{i,k} \cap C_{i+1,m}|}{|C_{i,k}|}$$

$C_{i,k}$ は文書集合 D_i の k 番目のクラスタ，$|C_{i,k}|$ は $C_{i,k}$ に属する文書断片の数を表す．一般に，集合間の関連度を求めるためには，以下の Jaccard 係数が用いられる．

$$Jaccard(C_{i,k}, C_{i+1,m}) = \frac{|C_{i,k} \cap C_{i+1,m}|}{|C_{i,k} \cup C_{i+1,m}|}$$

*2 http://chasen-legacy.sourceforge.jp/
*3 http://geta.ex.nii.ac.jp/

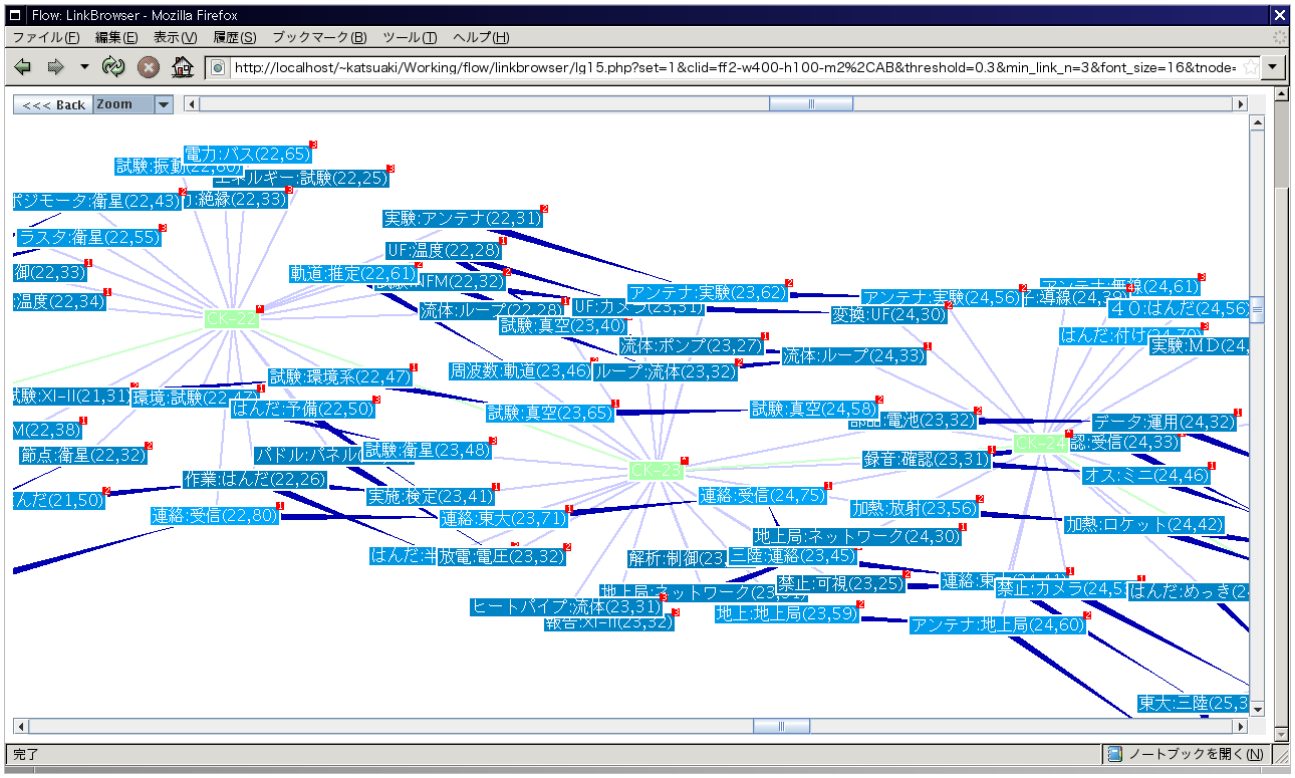

図 2 主題変化のグラフ表示例

しかし，$C_{i,k} \subseteq C_{i+1,m}$ となるとき，すなわち，$C_{i,k}$ が示す話題が他のもっと広い範囲の話題をもつ $C_{i+1,m}$ に統合されたとき，Jaccard 係数の分母の値が大きくなり，相関値が小さくなってしまう．そこで，上記の $sim(C_{i,k}, C_{i+1,m})$ を相関関数として用いた．

この $sim(C_{i,k}, C_{i+1,m})$ を用いて，$1 \leq i < N$，$1 \leq k, m \leq K$ の条件下で，各主題間の関連度の計算を行った．

§6 変化のグラフ化と表示

類似度関数 $sim(C_{i,k}, C_{i+1,m})$ により，隣接した文書集合，すなわち時間経過の上，直前直後の文書集合に属する主題間の関係を可視化する．表示には Touch Graph LinkBrowser*4を用いた．この際，類似度が閾値 T 以上の主題間にリンクを持たせることとし，主題 $C_{i,k}$，$C_{i+1,m}$ 間のリンク距離を，類似度の逆数と比例させ，以下のように定義した．

$$distance(C_{i,k}, C_{i+1,m}) \equiv -A \cdot sim(C_{i,k}, C_{i+1,m}) + B$$

各主題をグラフ上のひとつのノードとし，主題の特徴語と，文書集合の番号，および主題に含まれる断片の数を，ノードのラベルとして表示させた．特徴語は出現回数が多い単語を 2 つ選択した．表示例を図図 2 に示す．図図 2 の矢印の向きが，時間の経過方向を示す．

*4 http://sourceforge.net/projects/touchgraph

3·4 変化の再構成

§1 複数語による変化の再構成

主題構造の変化を俯瞰表示する以外に，ユーザが単語を指定することで，その単語に関連したノード（クラスタ）を表示させる機能を持たせた．グラフにおいて，クラスタ間の距離 $distance(C_{n,i}, C_{n+1,j})$ 以外に，単語が指定された回数以上出現するクラスタ $C_{n,i}$ を，単語を示すノードから，クラスタが属する文書集合 D_n の n に比例した距離を設定し表示を行う．これにより，ユーザがあらかじめ対象に関する知識を持っている場合，関連した複数の単語を指定することにより，それらの語の間の関連を把握することが可能である．

図図 3 に，CubeSat の設計過程で検討された複数の無線機を指定し，変化を再構成した例を示す．XI-IV の設計過程では，搭載無線機として汎用品である TEKKKS シリーズが設計初期に検討され，次に同じく汎用品の DJ シリーズが検討され，最終的に西無線による専用設計の無線機が搭載された．図図 3 からも，これらの無線機間の順序関係がわかる．DJ シリーズが西無線無線機の搭載が決定した後に再度取り上げられていることもわかるが，これは，衛星の機能試験のために DJ シリーズを用いたためであることが確認されている．

§2 特定語に関する変化の再構成

次に，主題の変化から，ユーザが注目している対象について，どのような作業が行われたかの一連のプロセスを抽出する．ユーザが対象を単語として指定するものとし，以下の処理を行う．

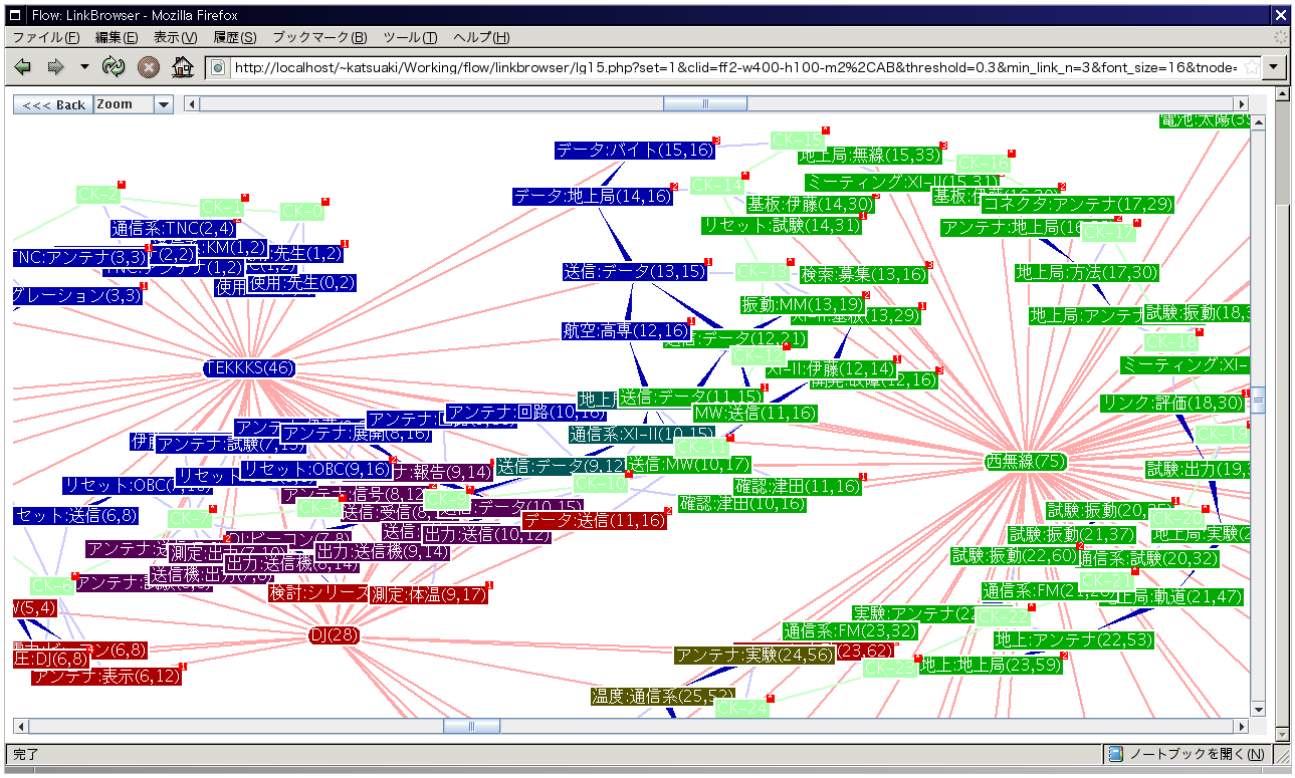

図 3 複数語による変化の再構成例

まず，ユーザに注目対象を示す語の記述を行わせ，これを含む主題を検索する．語と同時に，どの程度の回数，指定語が出現したら，対象に言及している主題と判断するかを判定する閾値の指定も行わせる．主題は文書断片の集合からなるので，文書断片のテキスト検索を行い，指定語の出現回数が閾値以上ならば，対象に言及した主題であると判断する．

続いて，対象に言及した主題から，言及内容を取得する．主題が含む文書断片のうち指定語を含むものを，新しい順に参照し，それらからサ変接続可能名詞（「〜する」「〜した」という記述が可能な名詞）を選択する．すなわち，対象を示す指定語に対する，主題内で最新の作業内容を示す名詞の抽出を行う．

対象に言及しているとして選択された主題を，取得した言及内容をラベルとして，同様により表示を行う．時間経過を示すノードとそのノード間のリンクを追加し，これらにそって選択された主題を表すノードを表示させた．

図図 4 に，“DJ” に関する主題変化を抽出した例を示す．DJ シリーズを用いて開発を行うこと，生産が中止されるため入手を急ぐこと，宇宙対策の検討，会計との関わりなど，議事録の記述内容が本稿で提案する手法により抽出されていることが確認できる．また，宇宙対策と購入が同じタイミングで検討されていることも，グラフ上で並行して表示されることにより，把握し易くなっている．

4. 例 2：個人の行動履歴

次に，個人の行動履歴から変化を同様に取り出し，グラフ化した結果から履歴に含まれる「非日常」状況の抽出を試みた例について述べる．

小型人工衛星の設計議事録を対象とした場合では，抽出対象とした文書群に含まれる内容について興味のあるユーザが，抽出した変化に注目して再構成を行った．一方，以下に述べる事例では，抽出対象の内容には特別な興味を持たない第三者が，グラフとして表現された変化からそこにあらわれる事象，とくにグラフ化された行動履歴から個人の「非日常」の認識を行うこと目的として，変化の抽出を行う．

4·1 マイ・ライフ・アシストサービス

マイ・ライフ・アシストサービス*5 は（株）NTT ドコモが平成 20 年度に経済産業省情報大航海プロジェクトの一環として実施したサービスである．サービス利用者の携帯電話に待ち受けアプリケーションを設定し，得られた情報をマイ・ライフ・サービス基盤へ集約，サービスプロバイダに対してこれをセキュアに提供する仕組みのもと，グルメ情報，口コミ情報などがサービス利用者に提供された．

*5 http://www.docomo.biz/html/member/mirai/013/

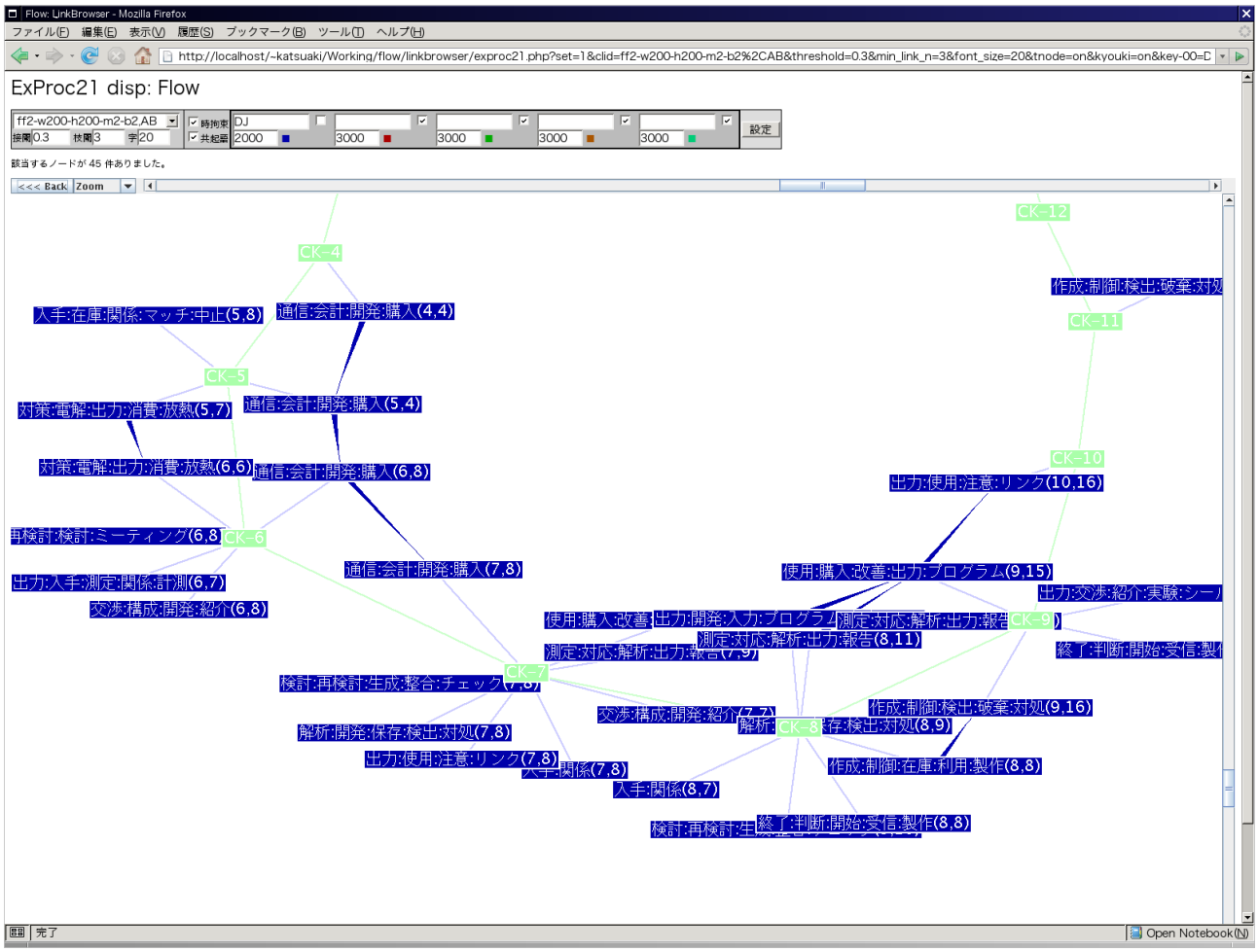

図 4 特定語に関する変化の再構成例

4·2 対象文書群

マイ・ライフ・サービス基盤に蓄積された情報のうち，利用者個人ごとの履歴を対象として3·3節節と同様の手法により変化のグラフ化を行い，利用者の「非日常」の抽出を試みた．蓄積された履歴には様々な種類があるが，本稿では，サービス中に利用者がどの画面ように画面遷移をさせたかを記録した操作履歴，飲食店舗の詳細や口コミの詳細など利用者が目にした文字列を記録したテキスト履歴を対象とする．

利用者の非日常は，グラフ化した変化のうち，時間経過をたどったときに新たに出現したノードに見て取れると考えられる．例えば，図図2において矢印により連結され，時間をまたがって長期間出現する主題は日常的な事項と考えられ，新たに出現する事項を非日常的な事項ととらえることができる．

4·3 変化の抽出

以下の手順により，個人のマイ・ライフ・アシストサービス利用履歴から，変化の抽出を行う．

(1) 履歴生成作成時刻による情報集合定義
(2) 履歴のクラスタリング
(3) クラスタの忘却処理
(4) クラスタ間の関連度計算とグラフ化

1，3，4では3·3節と同様の処理を行った．操作履歴は，操作に一定時間以上の間隔が空いたところを区切りとし，操作内容を単語とした単語列として文書を定義した．サービス利用者が目にしたテキストは，携帯電話を通したサービスであることもあり，短く単一の内容であったため，断片化を行わなかった．2のクラスタリングは，Probabilistic Latent Semantic Indexing（pLSI）[Hofmann 99] を用い縮約した結果の次元をクラスタとみなすことにより行った．得られたグラフの表示には Graphviz*6を用いた．なお，テキスト履歴の形態素解析には mecab*7を用いた．

テキスト履歴からの変化抽出例を図図5に示す．ノードの色は，ノードが示すクラスタ $C_{i,k}$ において，D_i で新たにあらわれた情報の割合，

$$\frac{|\{\mathbf{d}|E(D)+i\cdot S<c(\mathbf{d})\leq E(D)+(i+1)\cdot S\}|}{|C_{i,k}|}$$

に応じて水色に着色されるようにした．

*6 http://www.graphviz.org/
*7 http://mecab.sourceforge.net/

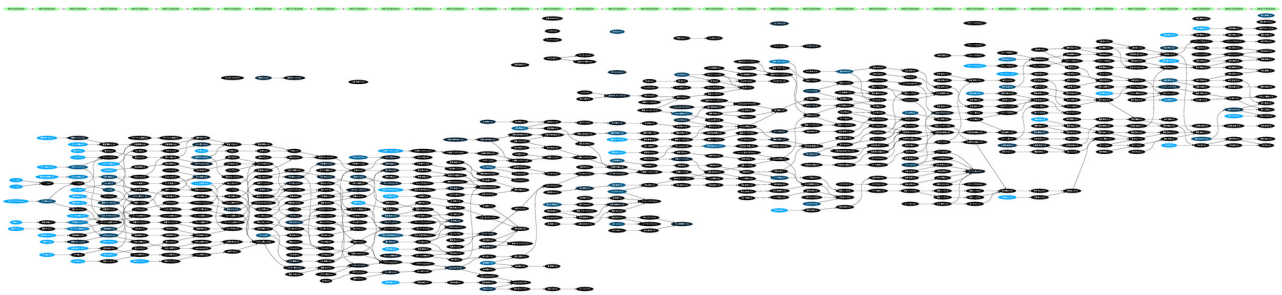

図 5　個人の行動履歴からの変化抽出例

4·4　利用者へのヒアリングとパラメータ調整

D_n を定義する時間間隔 S を 1 日とした上で，まず，クラスタ数 K を履歴数から生成されるクラスタのサイズを確認しながら決定し，続いて，忘却処理時の忘却割合 R，グラフ化の際のリンクを設定する閾値 T を変化させることにより，非日常と見なされる時間的に不連続なクラスタが出現する割合を調整した．

並行して，変化の抽出対象とした利用者に自身の履歴を確認しながらのヒアリングを行い，行動意図についてのコメントを取得し，そこから非日常と判定できるコメントと日付の特定を行った．

テキスト履歴では $K = 20$，$R = 0.5$，$T = 0.3$，操作履歴では $K = 10$，$R = 0.1$，$T = 0.5$ としたところ，得られた履歴の変化を表すグラフにおいて，不連続かつ新たな要素を多く含むクラスタ（非日常候補）の出現と，ヒアリング結果から判断された非日常事項との対応が確認できた．

また，サービスにおいて検索を積極的に行うなど能動的な利用者の場合にはテキスト履歴に，サービスが提示した情報が利用者自身の行動に適応しているか確認的に閲覧している利用者の場合は操作履歴に，それぞれ非日常が反映されやすいことなどが分かった．

5.　課　　題

3·4 節節にて述べた変化の再構成は，ユーザが指定する単語をもとにクラスタ内のテキスト検索を行うことにより実現している．例えば，ユーザが「TEKKKS」を指定すれば TEKKKS を含む変化が,「無線」を指定すれば無線を含む変化が抽出される.「無線」を指定した場合は「TEKKKS」のみを含むクラスタは再構成表示の対象として選択されない．図図 3 にあげたような設計途中の検討事項の変遷を抽出したいと考えるユーザが持つ意図は「無線機に関する設計過程の変化を把握したい」であることが多く,「TEKKKS と DJ と西無線に関する変化を知りたい」というところまで細かく指定できるためには，CubeSat XI-IV に関する背景知識が相当量必要であろう．

これを解決するために，変化の抽出結果と，部品間の概念データベースなどオントロジーに相当する情報を組み合わせることを検討している．これまでに，クラスタリングの階層性を用いて概念関係を抽出する検討を行った [田中 07] が，抽出された階層構造のラベリングにはやはりユーザの背景知識が必要である点には変わりがない．

また，4·4 節節では，変化をグラフ構造化した後，時間経過に沿って不連続に新規に出現する部分に着目して非日常の抽出を行っている．グラフ構造から情報を得る手法として，頻出する部分グラフを計算機による学習により見つけるグラフマイニング [鷲尾 06] が研究されている．これを用いることにより，抽出された変化の中に繰り返しあらわれるパターンを見つけ，何らかの意味づけを行うことも今後検討していく予定である．

6.　お　わ　り　に

本稿では，時間経過に沿って蓄積された情報からその変化の抽出を行い，ユーザの視点に基づいた再構成と，抽出された変化をグラフ化した形状からの変化状況の抽出について述べた．変化の抽出は，対象情報を生成時刻をもとに一定時間ごとの集合に分け，集合を分類し，類似した分類結果をリンク，あわせて古い情報を忘却させることにより，グラフ構造を得ることにより行った．次に，得られた変化に対し，対象情報の内容に興味があるユーザが，自身の興味を示す語により変化を再構成し表示させることを試みた．また，グラフにおいて不連続に出現するという特徴的な構造を持つ部分から，対象情報に記述が反映されていた何らかの非日常の抽出を試みた．

情報とそれが表す対象は，何者かの意図に基づく操作によって変化していく．さらに何者かがその変化を観察しその結果と自身の知識に基づき新たな意図を持ち，自身が持つまた別の知識に基づいて操作を行い，対象および情報を変化させる．この繰り返しにより，ある時点で観察される情報が形作られ，並行して蓄積する仕組みがあれば情報の蓄積も行われる．すなわち，情報は知識による何らかの変化の積み重ねであり，知識は何者かが何処かに変化を起こすために用いるものである．つまり，情報がある構造に至るまでに適用された知識は，その構造により直接は示されない．一方，その構造の時間経過に沿った変化が分かれば，使われた知識への手がかりとなるだろう．本稿では時間経過に沿って蓄積された情報か

ら，この変化の抽出と再構成を試みた．

また，変化そのものは知識ではないが，変化を再現することにより，さも知識を用いたかのような結果を，つくり出すことができるはずである．この際に，変化の組合せを，適用対象・文脈にあわせて再構成することにより，記録した変化からつくり出される「再現」の幅を広げ，よりそれらしい結果を生み出すことが可能だろう．

謝　　辞

本研究の一部は経済産業省平成 20 年度情報大航海プロジェクトにおける（株）NTT ドコモ「マイ・ライフ・アシストサービス」の一環として実施した．深く謝意を表する．

◇ 参　考　文　献 ◇

[Allan 02] Allan, J.: *Topic Detection and Tracking: Event-based Information Organization*, Kluwer Academic Publishers (2002)

[Havre 00] Havre, S., Hetzler, B., and Nowell, L.: ThemeRiver: Visualizing Theme Changes over Time, in *Proc. of IEEE Symposium on Information Visualization* (2000)

[Hearst 97] Hearst, M.: TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages, *Computational Linguistics*, Vol. 23, pp. 33–64 (1997)

[Hofmann 99] Hofmann, T.: Probabilistic latent semantic indexing, in *Proc. of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 50–57 (1999)

[Larson 99] Larson, W. J. and Wertz, J. R.: *Space Mission Analysis and Design, 3rd edition*, Microcosm Press (1999)

[Tanaka 04] Tanaka, K. and Takasu, A.: Topic Change Extraction from Problem-solving Records, in *Proc. of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics*, pp. 377–382 (2004)

[Tanaka 08] Tanaka, K., Akaishi, M., and Hori, K.: Reorganizing Topic Transitions in Design Process Records, in *Proc. of the Third International Conference on Knowledge, Information and Creativity Support Systems*, pp. 148–155 (2008)

[吉川 97] 吉川 弘之, 田浦 俊春, 小山 照夫, 伊藤 公俊：技術知の位相, 東京大学出版会 (1997)

[妻屋 02] 妻屋 彰, 永江 政郎, 若松 栄史, 白瀬 敬一, 荒井 栄司：共有と再利用を考慮した設計情報の統合表現, 人工知能学会論文誌, Vol. 17, pp. 53–60 (2002)

[松下 09] 松下 光範, 加藤 恒昭：情報編纂研究促進のための試み, 人工知能学会論文誌, Vol. 24, pp. 272–283 (2009)

[田中 07] 田中 克明, 赤石 美奈, , 堀 浩一：設計議事録からの主題構造変化の抽出と再構成, 第 65 回 人工知能学会人工知能基本問題研究会, pp. 29–34 (2007)

[武内 07] 武内 雅宇, 小路 悠介, 來村 徳信, 林 雄介, 池田 満, 溝口 理一郎：知識成長過程を指向した設計意図知識管理システムの構築, 人工知能学会論文誌, Vol. 22, pp. 263–275 (2007)

[鷲尾 06] 鷲尾 隆, 樋口 知之, 井元 清哉, 玉田 嘉紀, 佐藤 健, 元田 浩：グラフマイニングとその統計的モデリングへの応用, 統計数理, Vol. 54, pp. 315–331 (2006)