デジタル・ドキュメント 15－3
（1998. 11. 27）

情報流通向けテキストコンテンツ要約手法について

稲垣 博人　早川 和宏　田中 一男
NTT ヒューマンインタフェース研究所

デジタル化した情報が消費者に自由に行き来する情報流通社会において，種々の情報の流通・配信を効率的かつ効果的に行なうためのテキストコンテンツ要約手法について述べる．デジタル社会においては，ユーザが効率的に情報を取捨選択できるようにするため，ユーザの嗜好・リソース・状況に応じて最適なテキストコンテンツに変換して配信することが要求される．そのため，種々の要約タスクに応じて，最適な自然言語解析を用い，要約を生成する必要がある．本稿では，テキストコンテンツの要約手法として話題解析に基づく表層的な要約手法と，深層処理を利用するイベント解析に基づく要約手法を提案する．それぞれの手法によって生成される要約文の特徴や処理の特殊性などの要約特性について述べる．

Text contents summarization method for information circulation world

Hirohito Inagaki,　Kazuhiro Hayakawa, and　Kazuo Tanaka

NTT Human Interface Laboratories

In the information circulation world, text contents summarization is needed for effective and efficient circulation of information. To achieve effective and efficient communication between person, we have to maintain user preference, resource, and situation. There is not the only one best abstraction method, because of the difference of abstract characteristics in these abstract method. In this paper, we introduce two abstract methods. One is utilizing surface analysis method using topic in the document. The other is using deep semantic analysis based on event calculus in the document. Both abstraction methods have different characteristics to generate abstract, but it is necessary to merge both characteristics to gain better abstract.

1　はじめに

インターネットの普及，パソコン等の携帯端末・携帯電話の流行により，大量のデジタルのテキストコンテンツを自由に流通することができるようになった．これは，情報を効率良く流通する仕組みが大衆化した点によるものである．これらの情報流通におけるキーとなる要素は，以下である．

- 情報を入出力するハード・ソフト:PC，携帯端末
- 情報を発信するサービス:WWW，パソコン通信
- 情報をコミュニケートする機器: 携帯電話，電話
- 情報を発信するプロバイダ: 放送局，新聞社など

つまり，情報を入出力する場と，情報をサーキュレートする場の両方が揃うことにより，情報流通が効率的に行えるようになった．

一方で，今までの同一機器に対する同一情報の大量配信の時代から，個々人の嗜好，情報機器のリソース，状況などに合わせた少量配信という形で個の時代に変化しつつある．そのため，情報を提供する側としては，大衆

のメディアから，個のメディアに変換すべく，発信された情報を自動的に，再構成し，情報を配信する技術が必要とされている．

情報流通に向けたテキストコンテンツの配信では，個を対象とし，個のプリファレンスに応じて，最適な情報を編成し，配信することが要求される．つまり，ユーザに適したテキストコンテンツを配信するために，ユーザの嗜好，リソース，状況を考慮する必要がある．嗜好とは，ユーザがどのような情報を取得したいかを表す属性である．例えば，「A 社の情報がしりたい」，「B 業界の情報がしりたい」や，「A の B に対する対処を知りたい」，「M&A の状況について知りたい」など，明確に規定されるような事柄や，曖昧とした事柄など，種々ある．これらのユーザ嗜好に適したコンテンツを表示しなければ，ユーザにとっては，無駄な情報を表示してるだけに過ぎない．さらに，ユーザが使用するする機器のリソースに応じて（例えば，入出力速度，表示可能文字数，蓄積容量など）再編成し，最適な状況で配信しなければならない．そのため，テキストコンテンツを再編成する要約処理として以下のことが要求される．

- 適切な情報の要約（適切な内容表現）
- 適切な量での要約（適切な文字数での要約表現）

これらの要求を満足するような要約処理が，情報流通時代には要求される．そこで，本稿では，この情報流通時代に向けた要約処理を明確にするため，解析・編成・生成の各処理に分けて期待される要約処理手法について述べる．さらに，上記要約処理手法を実現するために表層処理による要約処理手法と，深層処理による要約処理手法の 2 種類の手法を提案するとともに，それぞれの特徴を情報流通の観点から考察する．

2　情報流通に向けた要約

要約処理は，大まかに以下の処理から構成される．

解析 テキストの自然言語解析

編成 テキストの重要部位抽出

生成 重要部位からの要約文生成

最適な情報を最適な量に再編成するような要約処理タスクを考えた場合，いかに適切に自然言語解析を行ない，テキストの重要部位を抽出できるかが重要となる．もちろん，要約文生成処理は，最終的な要約文を生成する上で重要であるが，コンテンツの中から重要部位を抽出することにより，適切な情報を適切な量に規定できるため，要約文生成処理に対する比重は比較的軽くなる．そこで，要約文生成においては，原文を基に要約文を生成することにより処理を単純化する．原文をベースとする要約生成では，以下の点でメリットがある．

- 文章の流れは原文を基にすればよい．
- 文章の用語は原文を基にすればよい．
- 不適切語句は発生しにくい．

通常の文生成では，文における主題，用語，文章の流れなどのすべてを加味して文生成を行なうような枠組が必要となる．原文をベースとすることにより，用語は原文と同様に統一され，かつ，生成する文の流れは，原文と同様に違和感のない文を生成することも可能となる．

テキストの自然言語解析では，表層解析を行なうか，深層解析を行なうかで，解析できる情報が異なる．表層解析の手法としては例えば，形態素解析，係り受け解析，キーワード抽出技術がある．深層解析であれば，意味解析，格解析などがある．一般に，表層解析の方が分野に依存せず，汎用性のある解析を行なうことが可能となる．一方，深層解析は分野に依存したり，特定の解析結果を生成することが多く，汎用性が失われる場合がある．

表層解析に求められるものは，文法的解析および，解析結果に基づく統語・統計的特徴抽出である．例えば，キーワード抽出を基にする要約処理では，統計的手法に基づいて，重要な語をチェックすることにより，文章中の重要部位を決定することができる[1), 2)]．さらに，表記の揺れ，コード上の揺れ（例えば，半角英語 ⇔ 全角英語，半角数字 ⇔ 全角数字など）の統一が期待される．ユーザ嗜好に基づき，重要部位を決定するためには，形態素の単語を利用し，ベクトル空間モデル的に，嗜好を表す単語のベクトルと要約文のベクトル間の演算に基づき，重要部位を計算することができる．また，係り受け解析に基づく共起関係もベクトル空間で表す手法[3)]ことにより，単語と同様に計算することが可能となる．もちろん，深層解析を行なっているわけではないため，決められた文字数で要約を生成することは難しい．どちらかというと，原文から文を抜き出す抄録となる可能性が高い．また，単にキーワードの統計的特徴抽出では，正確に重要な情報を捉えることが難しいため，何らかの修辞構造に基づきキーワードを取捨選択しなければならない．

一方，深層解析に求められるものとしては個々の単語の意味，意味的な同一性，文における役割，文章における役割などを求めることが要求される．例えば，文章構造に基づく要約手法[4]では，文章中の文間の関係を，“順接”，“逆接”，“例示”などで表し，文章が表すストーリに近い文を重要と規定し，要約として生成する．この場合，ユーザ嗜好は単語レベルというよりも，もっと高度な，たとえば，格関係レベル，文章構造レベルで求めることも可能となると同時に，ある程度決められた文字数で要約を生成することが可能となる．

もちろん，どのような要約タスクにおいても，深層解析に基づく要約処理が短時間に正確に行え，的確に要約を抽出することができれば，深層解析を用いた要約処理が最も適切な要約手法といえるが，実際は，深層解析は，ドメインが限られていたり，解析精度があるドメインであれば精度が高いが，他のドメインでは，解析精度が極度に低下してしまたり，実行時間がかかるなどの種々の制約が発生する場合がある．そのため，これらの制約に対して，ユーザ要求がマッチする場合には，深層解析を行なって要約を生成し，ユーザ要求にマッチしなかったり，深層解析が適切に実施できない場合には，表層解析を行なうことにより要約生成を行なう必要があると考える．

我々が提案する情報流通向けの要約手法は，表層的な要約手法では，キーワード抽出に基づく要約と文章修辞構造に基づく要約の両方を統合した手法である．これは，キーワード抽出では考慮できないキーワードの文章中の役割を修辞構造から抽出することにより，疑似的に文章構造に近い情報を求めているのである．つまり，文章における主題を意味するキーワードを取りだして要約を生成する手法である．一方，深層解析を利用した要約手法としては，イベント解析に基づく深層解析手法[5]を提案する．これは，文書に記述されている情報の中で，動作主体や，動作対象などのイベントに関連する情報を構造化することにより深層構造を抽出する手法である．格構造的なイベント構造を利用して要約を生成するため，イベント構造要素単位で重要部位を決定でき，種々の文字数での要約生成が可能となる．さらに，格構造的なイベント構造により，構造間でのマッチングによりユーザ嗜好をより高いレベルで整合性をとることが可能となる．図1に，それぞれ表層解析と深層解析に基づく要約処理の流れを示す．次に，提案する表層的要約手法と深層的要約手法のぞれぞれについて述べる．

3 表層解析による要約処理

表層解析による要約では，まず，入力文書の形態素解析を行ない，文を単語に分割し，個々の文に品詞を付与する．形態素解析が出力する各単語の品詞情報に基づき，文章中の話題を抽出し，話題を含む文または単文を重要部位とする．個々の文の重要度を基に，要約文として出力する候補を決定し，要約生成を行なう．

3.1 話題による文章構造化

ここで定義する話題とは，竹下ら[6]が定義した話題と同様なものである．つまり，話題とは「文章中で記述されている内容のうち，記述されている同じ内容をその文章中の言葉で表した語」であり，文書内のあるブロックの内容を明示する語又は名詞句相当語句が，個々のブロックにおける話題語となる．

3.2 快速覧による話題語を含む重要文抽出

話題の抽出には，テキスト向け話題抽出システム“快速覧”を利用した．快速覧システムは，対象となる文書の形態素解析情報，文書構造情報を用いて，文書中の話題を抽出するシステムである．形態素解析としては，InfoBee形態素解析エンジン[7]を用いている．快速覧は，話題構造を高精度で抽出することよりも，文書中のキーとなるような話題を抽出することが主であるため，複雑な話題構造を持たない．ここでは，話題の階層は，大局話題と局所話題の2階層とした．大局話題とは，明示的に表される比較的大きな話題である．一方，明示的ではないが，話題を転換したり，話題を展開したりする話題を局所話題と呼ぶ．局所話題は，大局話題の中にしか発生せず，大局話題中には複数の大局話題が入れ子で発生しうる．

大局話題は，以下の手順で決定される．

(1) 大局話題導入文の決定
「まず，」「第一に」などの語句（大局話題の導入手掛句）を含む文を大局話題導入文とする．

(2) 顕著名詞句の決定
以下のルールで名詞句（顕著名詞句）を抽出する．

- 大局話題導入文の中で「に関する」「について」などの話題を示す語句（明示マーカー）が後接続する名詞句
- 「は」「が」などの助詞（弱明示マーカー）が後接続する名詞句

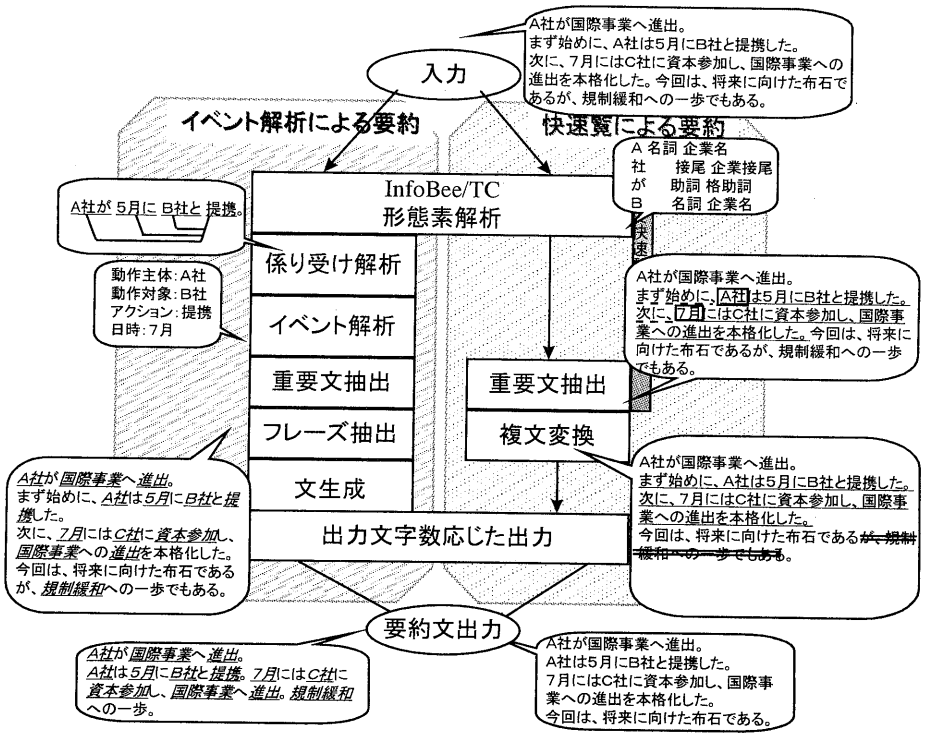

図 1: 要約処理の流れ

(3) 顕著名詞句の中で，話題の重要度や，品詞情報から最も重要な顕著名詞句を話題語とする．

局所話題における話題語は，大局話題の導入文でなく，かつ上記のような顕著名詞句のうち最も重要な顕著名詞句とした．

表 1に（非）明示マーカー，大局話題の導入手掛句，局所話題の導入手掛句の例を示す．局所話題の導入手掛句は，疑問表現や例示表現など，局所的に話題を変更するような表現が登録されている．

図 1の右側に快速覧による要約処理の流れの例を示した．図 1の文例では，“まず始めに”や，“次に”などの大局話題の導入手掛句があり，さらに，弱明示マーカーである“は”，“には”などがあるため，それぞれ，四角で囲った“A 社”と“7 月”が話題語として抽出される．また，これらの話題語を含む文（下線で示した文）が話題文として抽出される．

表 2に快速覧による話題抽出精度を示す．文章中で重要である部分を人間に抽出させ，快速覧が出力した話題文と比較した．文書の修辞構造を利用しているため，議事録などの修辞構造が適切に記述されている文書では非常に高い再現率を得ている．また，平均でも 86% と非

表 1: 話題抽出のための特殊表現

種類	特殊表現例
明示マーカー	とは，というのは，において
弱明示マーカー	では，が，を，である．には
大局話題の導入手掛句	まず，第一に，最初に，これに対しこの結果，次に，第二に，最後に
局所話題導入手掛句	例えば，1つに，その例としてとたずねる．と問われる

表 2: 文書タイプ別の話題抽出精度

文書タイプ	再現率	適合率
議事録	92%	57%
質問応答文	83%	57%
通知文書	77%	79%
平均	86%	62%

常に高い再現率を示しており，文書タイプによらず，汎用的に重要な部位を的確に抽出していることがわかる．

要約文生成処理では，抽出された話題を基に要約文候補を生成し，要求する文字数に応じて最適な要約文を選択・生成する．

要約文の候補として，以下のような候補が生成される．

(1) 話題文

(2) 話題文の表記変換処理

(3) 話題文の複文分割処理

(4) 話題単文

話題文とは，話題語を含む句点で終わる通常の文を意味する．話題文の表記変換処理では，意味的には全く同じであるが，記述文字数が少ない表現に変換する処理である．つまり，英数表記で全角であるものを半角に変換したり，単位表現や数値表現において原文より少ない文字数に変換する処理である．話題文の複文分割処理では，終止形で接続される複文を単文に分割し，話題語を含む単文を重要文とする．たとえば，”～するので，～した”などのように，終止形で終了する文の場合，理由・結果型の複文であることため，複文を構成するどちらの文も主語等の省略がなく，複文を分割しても，お互いの文が文として自立できる．図1の例では，“今回は，将来に向けた布石であるが，規制緩和の一歩でもある”の複文が，“今回は，将来に向けた布石である”と“規制緩和の一歩でもある”の文に分割され，話題語がある前の方の文が重要部位として抽出されている．話題単文とは，話題文の中で，話題語を含む読点で区切られた文字列のことを意味する．

つぎに，これらの要約文候補をもとに，ユーザ嗜好，リソースに合わせて，最適な要約を生成する．ユーザ嗜好がない一般的な場合では，要約文候補は以下の優先度に基づき候補が決定される．

(1) 大局話題を優先

(2) 話題文数を優先

(3) 同一話題文の他の要約文候補を優先

ユーザの嗜好を導入する場合は，各要約文候補とユーザ嗜好のキーワードによるベクトル空間モデルで類似度を比較することになる．ただし，この場合，話題は文章を弁別する語であるため，重要度は通常の単語より高めに設定することができる．

要求される文字数で要約を生成するために，以下のようなロジックで要約文候補を選択する．快速覧により抽出した話題文が要求される文字数以下であれば，抽出した要約文を原文の出現順に出力する．要求される文字数以上の場合，優先度の高い順に候補を出力する．この場合，要約の単位としては，文単位であるため，文より短い単位で処理できない．たとえば，新聞記事であれば，最も重要な文が平均 68.7 文字[†]で表現されており，100 文字程度の要約であれば，比較的文単位で要約を生成可能であるが，それ以上短い要約を生成する場合，適切な要約を生成することが難しくなる．

4 深層解析による要約処理

4.1 イベントによる文章構造化

深層解析では，イベント[8]という概念に基づいてテキストコンテンツを解析する．イベントとは，「世の中に発生する（した）事柄の変化」である．その事柄の変化を言葉で表現する要素がイベント要素である．必須のイベント要素としては，“動作主体”，“動作対象”，“アクション”，“場所（起点，終点）”，“時間（開始時刻，終了時刻）”などの 5 要素がある．さらに，これらの 5 要素を修飾するイベント修飾要素を規定した．

これらのイベント構造が人間にとっていかに，基本的であり，抽出されるエッセンスとしてもっとも適切であるかどうかを認知心理学的に求めた．文を被験者に一定時間見せ，その内容について自由に文を生成させる自由文生成テストを行った．被験者は 20 才から 30 才までの 18 人である．日刊工業新聞の記事を個々の段落ごとに 20 秒間呈示した．20 秒後には，記事を非呈示の状態とし，企業名だけは，継続的に呈示し，企業名＋“が”で始まる自由文を生成させた．約 10 記事について，生成

[†]この数値は，日刊工業新聞 93 年版から抽出したランダムに選んだ記事 50 件について，最も重要な文を被験者に抽出させた時の文の平均文字数である．

された自由文がイベント構造にどの程度整合するかを調べた．約93%の要素は，上記イベント構造で記述できることがわかった．呈示時間を40秒に増加させても，ほとんど記述される自由文に変化はなかった．これは，人間が認知する文の要素として，イベント構造は非常に高い整合性を持つことを意味している．

4.2 イベント解析による重要文抽出

イベント解析の処理では，まず，入力された文を形態素解析，係り受け解析した後，文を表現する木構造に変換する．木構造では，ノードとリンクで表現され，各ノードは単語を意味し，リンクは係り受け関係を意味する．個々の単語には，その単語の意味を表す意味素性や統語的制約を表す制約素性などの意味解析を行う上で必須の素性が記述される．木構造に基づき，与えられた素性を伝搬・統合することにより，意味解析が実施される．イベント解析によりイベントにおける動作主体やアクションなどのイベントを構成する要素が文から抽出できるが，これらのイベントを構成する要素だけでなく，各イベント要素を修飾する要素も同時に抽出される．イベントの修飾要素とは，事象要素に対して修飾関係にある表現や，「は」「と」「を」などの格助詞などの手掛かり表現を持つ要素である．これは，イベント要素が必ずしも文を構成する主要素（たとえば，主格や対格のような存在）でないからである．最終的にイベント構造の各要素には，要素に対応する単語や句などが記述される．イベント解析により抽出された要素を重要部位と判断し，さらに各イベントの優先度に基づき要約に必要なイベント情報を決定する．

4.2.1 イベント解析アルゴリズム

以下で例を用いてイベント解析のアルゴリズムについて説明する．「N社がA社の交換機を購入した」という文から表3のようなイベント構造が最終的に抽出される例を考えてみる．

イベント解析では，ロバストなイベントに基づく深層解析を行なうことを目的として設計されている．つまり，入力された文から何かしらのイベント情報を抽出することを目的としている．たとえば，イベント情報として，格解析のような文法格に基づく解析だけでなく，文法格以外の表現の中からもなるべく，イベント構造を抽出する．たとえば，極端な例では，名詞句の中に記述されているようなイベント構造も抽出される．これは，複合語のように名詞句が格的な構造を持つこともあるからである．また，ロバストな深層解析を目指すため，必ずしもすべての知識（意味素性など）が用意されていない状況でも解析を行えるようになっている．これは，それぞれの語がある意味素性などの属性をもち，それを他の語に伝播・統合するという基本的な働きを利用し，属性を持たない語に対してもその属性を推定するからである．たとえば，“社”という接尾語は，前の名詞が“会社”という属性を持つことを表しているのである．これらの素性の伝播・統合を基本アルゴリズムとした．そのため，必ずしも，“N社が”の“N”が会社であることを知識として持たなくとも処理を行なうことが可能となる．

表 3: イベント構造抽出例

要素名	パッセージ
動作主体	“N社が”
動作対象	“交換機を”
動作対象の修飾	“A社の”
アクション	“購入した”

このアルゴリズムによる企業名や商品名の推定精度を調べた．企業活動のドメインに絞って処理を行なった．企業活動に関する知識は，企業名等の固有名については，約8000語彙が用意され，企業の活動については，約1000語彙が用意されている．新聞記事約400件について，企業名については，57%の企業名については，“社”のようなレトリックにより推定され，トータルで91%の企業名を自動的に抽出できた．商品名などについても同様に84%まで自動的に抽出することができた．

動詞または，動詞同格の語句については，素性の伝播・統合ではなく，素性の制約の役割が強い．これは，ある意味では，格解析における格パターンと整合する部分であろう．たとえば，動詞“購入”は，企業分野では，動作主体として企業をもち，対象は，物品であることが制約として持つ．この制約を満たす必要がある．

このように，種々の語の素性を記述する知識と，その素性伝播の計算によりイベント構造が計算される．

以下で，例を示す．「N社がA社の交換機を購入した」という文は，形態素解析され，それぞれ，“N”，“社”，“が”という単語に分割されるとともに，個々の単語に品詞情報，素性情報が付与される．この場合，“N”や“A”には，企業という意味素性は付与されていない．“社”には，以下のような素性が付与される．表

記として，“社”があり，意味素性として，企業を表す属性を持つことが記述される．

“社”→

$$u1\begin{bmatrix} \text{表記} & = & \text{“社”} \\ \text{意味素性} & = & \text{企業} \end{bmatrix}$$

また，“交換機”には，物品という意味素性が付与される．

“交換機”→

$$u2\begin{bmatrix} \text{表記} & = & \text{“交換機”} \\ \text{意味素性} & = & \text{物品} \end{bmatrix}$$

一方，この文の動詞である“購入”は，以下のような素性を持つ．

“購入”→

$$u3\begin{bmatrix} \text{表記} = \text{“購入”} \\ \text{意味素性} = \text{企業活動} \\ \text{制約} = \begin{bmatrix} \text{動作主体} = \text{企業} \\ \text{アクション} = \text{企業活動} \\ \text{動作対象} = \text{物品} \\ \text{動作時間} = \text{時間} \\ \cdots = \cdots \end{bmatrix} \end{bmatrix}$$

制約と記述された部分が，イベント解析における格情報にあたる部分である．この制約に整合する要素が文の木構造から算出される．最終的に木構造の一番上のノードには，制約の素性を満たし，かつ，素性の伝播と統合により計算された意味解析結果が記録される．図2に例文の意味解析結果を示す．

このような素性に基づくイベント構造の計算により，動作主体，動作対象などが明確化される．この手法に基づく解析では，人間が作成した正解に対して，45%の精度で自動的に各要素を抽出することができた．さらに，複数の被験者の抽出したイベント構造との再現率，適合率を求めた．適合率は，被験者が自由文で作成したイベント構造を用いて計算すると72%であった．さらに，再現率については，新聞記事の1段落を被験者に呈示した後，計算機が出力した結果を呈示し，計算機が出力した結果が適切に表現しているかどうかを被験者にYESとNOで解答してもらうことにより算出した．その結果，再現率は60%であることがわかった．再現率，適合率共に高い精度で深層解析が行えることがわかった．

4.2.2 重要部位抽出

深層解析に基づく要約では，イベント解析のドメイン依存性とイベント構造による情報のエッセンス化を利用する．つまり，企業の活動なら，企業の活動に関するイベント構造だけを文中から抽出することができるため，一種の情報フィルタリングとして本イベント解析を利用することができる．また，先の認知心理学テストでもわかるとおり，文に記述されている情報すべての構造化するわけではなく，人間が認知するもっとも基本的かつ重要な情報だけをイベント構造として抽出しているため，情報のエッセンスだけが文中から抽出される．このドメインに関するフィルタリングとイベント構造に基づく構造化により，文章中からあるドメインに関するエッセンス情報を抽出することが可能となる．たとえば，要約として不要な例示表現や，状況説明などの種々の曖昧な表現はすべてイベント構造には当てはまらないため，重要とは判断されない．

4.2.3 要約生成

イベント構造解析により抽出されたイベント情報を組み合わせて要約を生成することが可能であるが，ここで，文章中から抽出されたイベント情報の優先度を付与する．たとえば，通常の新聞記事では，ほとんどの場合，先頭の文から重要な文を記述している場合が多く[‡]，その意味では，新聞記事などでは，記事の文の順にイベント情報の優先度をつけることも考えられる．また，マスメディアなどの多くのメディアで重要であると判断された情報こそ多くの人が重要であると判断するという仮定のもとイベントの重要度を決定する手法[10]もある．さらに，ユーザ嗜好に応じて，これらの優先度に対して増減をする必要がある．この場合は，イベント構造とユーザ嗜好との各スロットの整合性から判断する方法がとれる．

要求された文字数で要約を生成する場合，深層解析では，意味的に単語単位で扱っているため，高度な取捨選択が可能となる．個々のイベント構造において以下のように優先度付けすることが可能である．

- 最少の3イベント要素（動作主体，動作，対象）
- 基本5イベント要素（動作主体，動作，対象，時間，場所）
- 5イベント修飾要素

これらの優先度に応じて要約に最適なイベント要素を決

[‡]試験的に，50件の新聞記事について100文字の要約を被験者に作成させたところ，被験者が抽出した要約の96%が新聞記事の最初から順に抽出されていた．

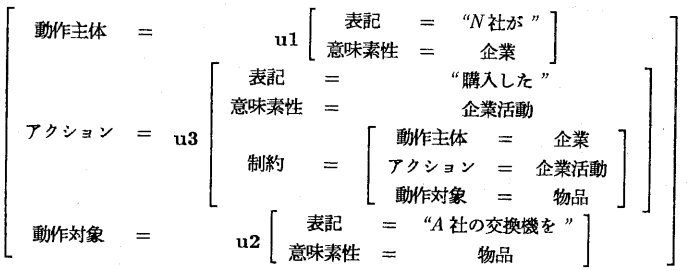

図 2: 木構造の計算結果

定し，意味的にも，文脈的にも適切となるように，必要な語句を用いて表現する．

5 まとめ

情報流通社会に向け，ユーザの嗜好，リソース，状況を考慮し，最適な情報を最適な量で再編成する要約処理について述べた．特に，表層処理では，話題に基づく要約手法を提案し，深層解析では，イベント解析を用いた要約手法を提案した．話題に基づく要約手法では，形態素解析と快速覧解析とにより分野に依存せず解析ができ，かつ，再現率 86% の精度で，重要部位を抽出することができた．一方，イベント解析を用いた深層処理に基づく要約手法では，分野に依存した処理となるため，分野ごとに知識を構築する必要性があるなどの繁雑な面もあるが，イベント構造は，人間が認知する構造にマッチしているため，高い精度で重要な部位を抽出できると共に，要求される文字数にマッチした要約文を生成しやすい．表層解析と深層解析のそれぞれは，一長一短があり，どちらが決定的な要約手法であるとは言えない．その意味では，これらの表層解析と深層解析のそれぞれを利用して，ユーザ要求に最も適合し，適切な要約を生成できる手法を状況に応じて選択する必要がある．

参 考 文 献

1) H. P. Luhn. The automatic creation of literature abstract. *IBM Journal, Vol2*, 1958.

2) 鈴木康広, 上窪真一, 栃内香次, 永田邦一. 高頻度隣接語を利用した科学技術文献の自動抄録. 情報処理学会第 32 回全国大会, 4T-11, 1986.

3) 森辰則, 大森信行, 内間圭介, 岡村淳, 中川裕志. 電子化マニュアルにおける自動ハイパーテキスト化手法. 情報処理学会デジタルドキュメント研究会,DD9-3, 1997.

4) 内海功朗, 重永実. 英語文章の大意生成. 情報処理学会自然言語研究会, NL54-8, 1986.

5) Hirohito Inagaki and Tohru Nakagawa. An abstraction method using a semantic engine based on language information structure. *Coling-92*, 1992.

6) 竹下敦, 井上孝史, 田中一男. テキストの概要把握支援のための話題構造抽出. 情報処理, 1996.

7) 井上孝史, 大久保雅且, 杉崎正之. Infobee テキスト検索情報検索技術. NTT R&D ジャーナル,Vol.46,No.10,pp.1103-1108, 1997.

8) 稲垣博人. 事象解析による要約情報の抽出. 情報処理学会自然言語研究会,NL84-3, 1991.

9) 稲垣博人, 中川透. 出来事型情報の構造化. 情報処理学会第 46 回全国大会,4A-7, 1993.

10) 稲垣博人, 早川和宏, 田中一男. 類似意味内容の統合による伝達型電子化文書要約方式の提案. 情報処理学会第 57 回全国大会,4R-11, 1998.