データベースシステム 111－10
（1997. 1. 22）

時刻印付ノードリンクグラフによるビデオ映像のデータベース化

是津耕司† 上原邦昭†† 田中克己††† 木邑信夫†

†通信・放送機構 神戸リサーチセンター
††神戸大学工学部情報知能工学科
†††神戸大学大学院自然科学研究科知能科学専攻

本論文では、" 時刻印付ノードリンクグラフ" によるビデオ映像記述方法を提案する。時刻印付ノードリンクグラフによるビデオ映像記述は、ビデオ映像から連想される言葉や印象をキーワードに選び、これらキーワード間で関連があると思われるものどうしをリンクで結んでグラフを形成する手法である。各キーワードには、それが指し示す映像部分の時刻をつけ、ビデオ映像の時間的属性を表現している。本記述方式により、これまでは取り扱うことが難しかったストーリー性の薄い映像の記述・検索がうまく行なえることを示す。最後に、本記述方式を東映アニメ予告編映像に適用した結果をもとに考察を加える。

A Time-stamped Node-link Graph for Video databases

Kohji Zettsu† Kuniaki Ueharа†† Katsumi Tanaka††† Nobuo Kimura†

†Kobe Research Center, TAO
††Department of Computer and Systems Engineering,
Faculty of Engineering, Kobe University
†††Division of Intelligence Science,
Graduate School of Science and Technology, Kobe University

This papaer describes the video description method using "Time-stamped Node-link Graph". By Time-stamped Node-link Graph, a video is described as the graph that consists of keywords and relations of them, that become nodes and links of graph respectively. Those keywords represent author's impressions of the video. Then, he links those keywords when he finds some relations between them. Each keyword is marked with the time when it occuers to author. (so, we call it "time-stamped") This paper shows Time-stamped Node-link Graph can handle those videos properly that don't have enough story in them and have been difficult to be described so far. We use the animation video previews created by TOEI film studio as example.

1 はじめに

ビデオ映像をデータベース化する際には、それぞれの映像に対して検索のためのインデックス付けを行なう必要があるが、これには大きく分けて２つのアプローチがある。１つは、映像の信号情報から得られる特徴量（１次情報）をインデックスとして利用する方法と、もう１つは、キーワードなど人間が介在して外部から映像に付ける情報（２次情報）をインデックスとして利用する方法である。どちらの方法を採用するかは、どのような検索インタフェースを想定するかによるが、現状ではより親しみやすい文字列で検索要求を表現して検索をすることが多い。この場合、映像に対するインデックスは２次情報を使ったものになる。

２次情報によるインデックスづけを行なう場合、これまでは主にシーンの情景やストーリーにもとづいた映像記述からインデックスを得ていた。しかし、シーンがめまぐるしく変化してストーリー性が乏しい映像、例えば今回研究素材として利用した予告編映像などは、情景やストーリーにもとづいた映像記述が難しく、これらをインデックスづけし検索することは困難であった。

そこで本研究では、映像の内容に対する断片的な記述、すなわち映像のそれぞれのシーンに対する印象や感想とそれらの関連性をもとに映像のインデックスづけを行なう方法として「時刻印付ノードリンクグラフ」を提案する。本論文では、時刻印付ノードリンクグラフによる映像記述の特徴と、これを用いたビデオ映像の記述および検索について述べる。

2 記述の特徴

これまでに提案されたビデオ映像記述方式は、キーワード [1] やレコード [2] あるいはグラフ構造 [3][4] により特定の時間区間に上映されるシーンの情景やストーリーを記述し、これを時間軸上に並べることで映像全体を記述していた。これらの方法では、求めるシーンを検索するために各時間区間における映像シーンを正確に記述することが要求されるが、これは非常に困難な作業である。

困難と感じるのは、記述者である我々自身が同じ映像シーンに対してさまざまな印象や感想を持ち、それらを拠り所に映像に対する説明を与えようとするからである。我々が映像を理解する際には、その前に流れた映像との因果関係や、論評や宣伝など映像以外から得た情報を使いながら理解することが多い。これまでの記述方法のように、時間で区切られた映像部分の記述を時間軸上に並べるだけでは、こうした我々の映像理解にもとづいた記述を行なうことはことは難しい。

時刻印付ノードリンクグラフでは、厳密なシーンの説明によって映像を記述するのではなく、映像に対する印象や感想あるいは映像に付帯する情報とこれらの間の関連を使いながら、内容の意味的なまとまりやつながりをもとに映像を記述し検索することを目指している。したがって、時刻印付ノードリンクグラフによる映像記述は、我々人間の映像理解により近い記述方法と言える。

3 ビデオ映像の記述

本節では、時刻印付ノードリンクグラフによるビデオ映像の記述方法について述べる。

3.1 記述手順

1. キーワード付け

時刻印付ノードリンクグラフによるビデオ映像記述では、記述者が映像を見て思いついた印象や感想を記述者の言葉で表現し、これを映像に対するキーワードとして付与する。キーワードに使う言葉は、自然言語による自由な文 (sentence) である。このキーワードは、映像のある部分を見た時に思いついたはずなので、その思いついた時刻（映像のタイム・コード）をキーワードに付ける。これが「時刻印付」と呼ぶ所以である。この時刻印のつけられたキーワードを時刻印付ノードリンクグラフではノードと呼び、$n = (kwd(n), t(n))$ と表す。ここで、$kwd(n)$ はノード n におけるキーワードであり、$t(n)$ はその時刻である。また、すべてのノードの集合を $N = \{n_i\}$ と表す。

映像に対する印象や感想だけではなく、ビデオ映像に付帯する情報（例えば製作スタッフ、俳優、公開日など）からもキーワード付けを行なう。これらのキーワードには映像全体にあてはまるものが多く、特定の時刻をつけることができないものもある。こうした時刻を付けられないキーワードを持つノードは、時刻印の値のない “時刻なしノード” として扱う。

2. リンク付け

次に、記述者がノード n_i とノード n_j $(n_i, n_j \in N)$ に付けられたそれぞれのキーワード間に何らかの内容的な関連があると思った場合、これらの間をリンクでつなぎ相互に関連があることを示す。ノード n_i と n_j 間のリンクを $\ell = (n_i, n_j)$ と表す。リンクは関連の有無を示すだけで、関連の具体的な内容は示さない。ただし、その関連が見い出された状況に応じて、リンクを次のように分類する。

通常リンク 記述者が見た映像の内容だけにもとづいて付けられたリンク

常識リンク ビデオ映像の中には直接出てこないが、記述者が既に持っている知識や常識をもとに関連付けたリンク

汎化リンク 一方のキーワードが他方のキーワードのことをより抽象的な言葉で言い換えているという関係を表すリンク

このようにして得られたすべてのリンクの集合を $L = \{\ell_i\}$ とする。また、ノード集合 N とリンク集合 L を使って、時刻印付ノードリンクグラフを $G = (N, L)$ と表す。

3.2 記述の例

時刻印付ノードリンクグラフによる記述の例として、東映アニメ予告編映像３３作品の中から「サイボーグ 009」（1966 年 東映動画スタジオ製作、3 分 18 秒）(図 1) を記述した結果を図 2 に示す。図 2 では、通常の（時刻のついた）ノードをタイムチャー

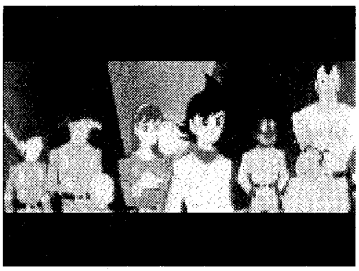


図 1: 東映アニメ予告編映像「サイボーグ 009」

ト内に、時刻なしノードをタイムチャートの枠外に記述している。

3.3 考察

「サイボーグ 009」の例において、ノード ("ずっと進んだサイボーグ", 00:00:37:26) から放射状にリンクが張られているが、これはこのノードが付けられたシーンの内容 (「君はすごいサイボーグなんだ」と決定的に言う) に対して、後に続く数シーンで具体的な説明 (どうすごいサイボーグなのか。("マッハ 5 で空を飛ぶ 002", 00:01:05:04) 、("20 万馬力の 005, 00:01:20:10) など) をしている。また同様に、ノード ("怪獣が攻撃", 00:00:17:11) がノード ("恐竜ロボット", -:-:-:-) を介してノード ("009 怪獣に向かっていく", 00:02:36:24) と結びつけられていることから、これら離れた 2 つのノードに現れる "怪獣" は同じ "恐竜ロボット" のことを言っているのだとわかる。このように、時刻印付ノードリンクグラフでは、時間的に離れていても意味的にはつながっている映像間の関係をリンクによって表現することができる。従来の記述方法のように各シーンに対する記述を時間軸に沿って並べて記述するだけでは、このような関係を表現することはできない。

4 ビデオ映像の検索

本節では、時刻印付ノードリンクグラフを用いたビデオ映像の検索について述べる。検索の目的は、検索要求として自然言語による文を与え、検索文に含まれる単語間の関連性からこの検索要求を最もよく表現できるカットを選び出すことである。

4.1 検索手順

1. キーワード・マッチング

キーワード・マッチングとは、検索文に含まれる各単語 (検索語) にマッチするノードを見つけ出すことである。検索文 Q に含まれる検索語 q_i にマッチしたノードを検索語 q_i に対するマッチノード $m(q_i)$ として定義する。

マッチノードの検索は以下のようにして行なわれる。

1. 各ノードに付けられたキーワードと検索文との間で単語単位の照合を行なう。その際、検索文に含まれる単語の同義語についても照合を行なう。

2. 常識リンクおよび汎化リンクを使ったマッチングを行なう。

   - ある検索語に対するマッチノードと常識リンクで結ばれたノードもこの検索語に対するマッチノードとみなす。
   - 汎化リンクで結ばれたノードでは、親ノードがマッチした場合その子ノードもマッチしたとみなす。

マッチングによって得られたマッチノードを各検索語ごとに組合せ、検索文 Q に対するマッチノードベクトル $\mathbf{m}(Q) = (m_{i_1}(q_1), m_{i_2}(q_2), \cdots, m_{i_n}(q_n))$ ($m_{i_k}(q_k)$ は検索語 q_k に対する任意のマッチノード) を作る。これを可能なすべてのマッチノードの組合せについて作成し、検索文に対するマッチノード集合 $\{\mathbf{m}_i(Q)\}$ を得る。

2. グラフの検索

次に、各マッチノードベクトルについて、マッチノードベクトルの成分にあるマッチノードをすべて含む最小の部分グラフを記述グラフの中から探し出す。これは、マッチノードおよびマッチノード間を結ぶ際に経由する中継ノードに対する最小全域木 (minimum spanning tree) を見つけ出すことである。現在は、リンクの重みはすべて 1 としている。こうしてマッチノードベクトル $\mathbf{m}_i(Q)$ に対して選ばれた最小部分グラフ $s_i(Q) = (N_{s_i(Q)}, L_{s_i(Q)})$ ($N_{s_i(Q)}$, $L_{s_i(Q)}$ はそれぞれ $s_i(Q)$ のノード集合とリンク集合) は、$\mathbf{m}_i(Q)$ に含まれるマッチノードに付けられたキーワードを使った時に検索文 Q に含ま

れる単語間の関連を最もよく表している記述部分であると言える。このグラフ検索を、マッチノード集合に含まれるすべてのマッチノードベクトルについておこない、検索文 Q に対する最小部分グラフ集合 $S(Q) = \{s_i(Q)\}$ を得る。$s_i(Q)$ のうち、重み合計が小さいものほど検索語の関連をよく表わしていることになる。

3. カットの選択

最後に、グラフ検索によって選び出された部分グラフに対応する映像カットをビデオ映像から選び出す。これは、$s_i(Q)$ に含まれるすべてのノード $n_j^{s_i(Q)} \in N_{s_i(Q)}$ について、各ノードに付けられた時刻に上映されるビデオ映像をカット単位で抜き出してくることである。ただし、$n_j^{s_i(Q)}$ が時刻なしノードの場合は、このノードに対するカットの選択は行なわない。こうして得られた $s_i(Q)$ に対するカットの集合が1つの検索結果となる。この操作を $S(Q)$ に含まれるすべての $s_i(Q)$ に対して行うことで、検索文 Q に対するすべての検索結果映像を得ることができる。

4.2 検索の例

時刻印付ノードリンクグラフによる検索の例として、同じく「サイボーグ 009」の記述に対して「サイボーグが敵と戦っている」シーンを検索した結果を図2に示す。図2では、マッチノードおよび各マッチノードベクトルを含む最小部分グラフを記述グラフ上に、また各最小部分グラフから検索結果として選ばれたカットを対応する時間区間で表している。さらに、我々が見て検索要求に合致していると思われるシーンを正解シーンとして併記してある。

4.3 考察

「サイボーグ 009」による検索の例では、検索結果として選ばれたカットのほぼすべてが該当する映像部分（正解シーン）に含まれ、時刻印付ノードリンクグラフによる映像検索が可能であることが示された。正解シーンのうち検索結果として選ばれなかったものもあるが、これは現段階ではまだこの作品に対する記述量が不足していることが原因である。

その一方で、検索結果の中にはカットがとびとびに選ばれてしまい、得られた映像が意味のないものになってしまったものもある。これは、ノードに付けられた時刻とこの時刻に上映されるカットが必ずしも1対1に対応しないことが原因と考えられる。ノードに付けられた時刻はそのキーワードを「思いついた」時刻であり、キーワードにはこの時刻より前に上映された映像の内容や外部から得た情報も含んでいる。したがって、ノードはその時刻の近辺にある映像に対応づけられると考えることができる。そのため、従来の記述方法のように、記述に表わされた対応だけから単純に特定のキーワードに対する連続した映像を選択することが難しい。グラフ検索後の映像選択の際に何らかの処理を施す必要があるだろう。

5 映像記述のクラスター化

本節では、時刻印付ノードリンクグラフのクラスター化について述べる。クラスター化の目的は、時刻印付ノードリンクグラフの中から関連の密度をもとに「まとまり」を見つけ出すことであり、これはビデオ映像の中の意味的なまとまりを発見することにつながる。

5.1 クラスター化手順

ここでは、純粋に映像内容からつけられた関連だけをもとにクラスタリングを行なうので、常識リンクおよび汎化リンクはクラスリングする記述グラフから除いておく。

1. 時刻印付ノードリンクグラフに含まれる連結グラフをそれぞれクラスターとする。
2. 各クラスターに含まれる連結グラフの橋 (bridge)[1] を見つけ出し、橋を境にクラスターを2つに分割する。分割されたクラスター間をこの橋で結ぶ。

   ただし、橋になったリンクが端点につながるリンクであったり、あるいはこのリンクでの分割ができないと事前に定義されている（常識的に関連を断ち切ることができないリンクなど）場合は、そのリンクでの分割は行なわない。また、1つのクラスターに複数の橋が存在する場合は、橋の重さが最小のものを選んでここで分割する。

[1] このリンクを除くと、もとのグラフが非連結になるリンク。

3. 各クラスターに含まれる連結グラフのカット点(cut-vertex)[2]を見つけ出し、カット点を境にクラスターを2つに分割する。このカット点は分割された双方のクラスターに含まれ、双方のクラスターはカット点どうしの間で重さ0のリンクで結ばれる。

以上の手続きを、これ以上クラスターが分割できなくなるまで繰り返す。

5.2 クラスター化の例

時刻印付ノードリンクグラフによるクラスター化の例として、同じく「サイボーグ009」の記述に対してクラスター化を行った結果を図3に示す。図3では、各クラスターを記述グラフ上の各領域として表している。

5.3 考察

クラスター化の結果得られた各クラスター領域中のキーワードを総合してみると、それぞれのクラスターが内容的におおよそのかたまりを構成していることが分かる。このことから、時刻印付ノードリンクグラフのクラスター化によって、映像の意味的なまとまりを見い出すことが可能であると言える。ただし、そのためにはクラスター化の際に分断できない（常識的に関連を切れない）リンクを慎重に選択しておく必要がある。

6 今後の課題

今回提案した時刻印付ノードリンクグラフによるビデオ映像の記述を通して、気づいた点および今後の課題について述べる。

リンクの自動生成

キーワード間のリンクの設定は、本来すべて記述者の意志によって行なわれるべきものであるが、実際すべてを記述者が行なうのは大変な作業であり、基本的な関連を見落としてしまったために検索で全く期待外の結果を出してしまう恐れがある。そこで記述者のリンク付けを支援する目的で、自明な関連には自動的にリンクを張るようにする。例えば、一部の常識リンクや汎化リンクなどがこれにあてはまる。これらは辞書やシソーラスからある程度機械的にリンクを生成することができる。

キーワードのベクトル化

現在は自然言語文として表現しているキーワードを、ベクトルによって表現することも考えられる。例えば、ベクトルを (<サイボーグ>, <怪獣>, <戦闘機>, <戦う>, <助ける>, <怒る>) と定義すると、「サイボーグが怪獣と戦っている」は、(1, 1, 0, 1, 0, 0) と表現できる。キーワードをベクトル化することにより、次のような効果が期待される。

- キーワード間の類似度をベクトル間の距離で表現し、検索の際のキーワード・マッチングやグラフ探索、あるいはクラスター化の評価基準として利用する。
- 同じベクトル成分にビットが立てば、これらのキーワード間に関連があることを記述者が明示的にリンクを張らなくても表すことができる。

キーワードのベクトル化の大きな問題は、記述者が自然言語で表記したキーワードをどのようにベクトル化すべきかということである。

不連続部分の評価

検索の考察でも述べたように、検索の結果時間的に不連続な映像が選び出されてしまう場合がある。したがって、検索結果にこうした時間的な不連続が生じているかどうかを判断し適切な映像を補間する処理が必要である。

またこれとは逆に、選ばれた映像の中に、検索要求とは全く関係のない、本来なら入るべきでないカットが含まれている場合もある（内容の不連続。例えばCMなど）。こうした映像を検索結果から除去する処理も必要である。

いずれの場合も、時間的に近いものほど内容の関連性は強いという映像の一般的性質から、選択された映像とこれらの近傍の映像との類似度を評価基準として映像の補間／除去を行なうことが考えられる。

リンクの重み付け

前述したように、ビデオ映像においては一般的に時間的に近くにある映像ほどより強い関連をもつと

[2]この点を除くと、もとのグラフが非連結になる点

考えられる。したがってノード間の時間的な隔たりの大きさにしたがってリンクの重み付けをすることで、検索結果として時間的にまとまったものが選び出されやすくなり、前述した映像の不連続が起こりにくくなると考えられる。

キーワードの補間

記述者が自由な表現で書き出したキーワードの中には、主語や目的語、動詞などが抜けているため何を言っているのか分からないものもある。例えば「敵と戦う」(誰が?) や「009」(がどうした?) などである。こうしたキーワードが多く含まれていると、記述内容が全般的に隙間の多い (sparse な) ものになってしまい、検索結果の映像に不連続を生み出す恐れがある。そこで、ノード間に張られたリンクをたどり、キーワードに欠けた言葉を補間しながらキーワードを集約していくことで、記述内容をより密度の濃いものにしていくことが考えられる。

クラスターの意味づけ

今回のクラスター化では、時刻印付ノードリンクグラフによる記述から関連性にもとづいて意味的なまとまりを発見することができたが、まだそれぞれのクラスターへの意味づけをおこなっていない。クラスターに意味づけをする方法には次のようなものが考えられる。

- 前述したキーワードのベクトル化を利用して、クラスターに含まれるノードのキーワード・ベクトル間の演算から新たにクラスターのキーワード・ベクトルを作り出す。
- 前述したキーワードの補間を利用して、クラスターに含まれるグラフのリンクをたどり、クラスターを表現する文の主語や動詞を補間ながらキーワードを組み立てていく。

複数の作品間にわたる記述と検索

今回は東映アニメ予告編映像の1作品を対象に記述と検索を説明した。今後はさらに多くの作品を対象に、特定の作品の特定の映像シーンを選び出せるよう、複数の作品間にわたる記述と検索へと発展させていく。

7 おわりに

今回の時刻印付ノードリンクグラフによるビデオ映像の記述では、映像に対する断片的な記述とこれらの関連によって映像にインデックスを付けることで、キーワードの関連性にもとづいて映像シーンを検索することができることが示された。そして、本記述方式が予告編映像のようなストーリー性の薄いビデオ映像に対しても有効に働くことを示した。本方式をさらに有用なものにするためにはまだ多くの課題が残されているが、本方式は映像を自然言語で記述し検索するための基本的かつ有効なアプローチとなり得るであろう。

謝辞

本研究のために貴重な映像資料を提供していただいた、東映株式会社様に感謝致します。本研究は、一部、文部省科学研究費重点領域研究 (課題番号 08244103) による。

参考文献

[1] 柴田 正啓: 映像の内容記述モデルとその映像構造化への応用, 電子情報通信学会論文誌 D-II, Vol.J78-D-II, No.5, pp754-764 (1995)

[2] E.Oomoto, K.Tanaka: OVID: Design and Implementation of a Video-Object Database System, IEEE Transactions on Knowledge and Data Engineering, Vol.5 No.4, pages 629-643 (1993)

[3] 前原 恵太, 福田 慶朗, 上原 邦昭: グラフ表現による動画像内容記述からの階層的場面構造の構築, 情報処理学会研究会報告, 96-DBS-109, Vol.96 No.68 (1996)

[4] 柴田 正啓: デスクトップ映像製作, ADBS '96 (1996)

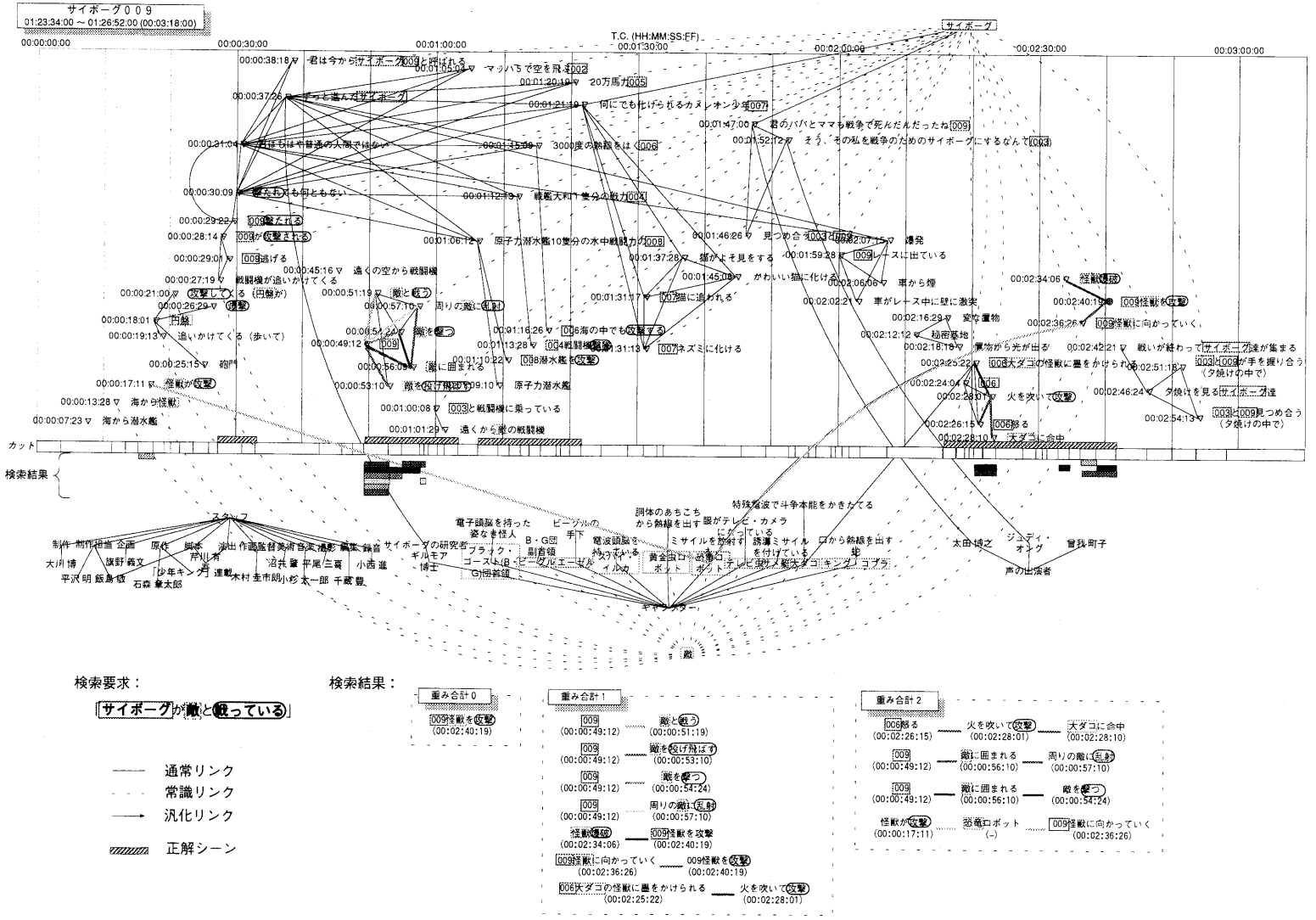
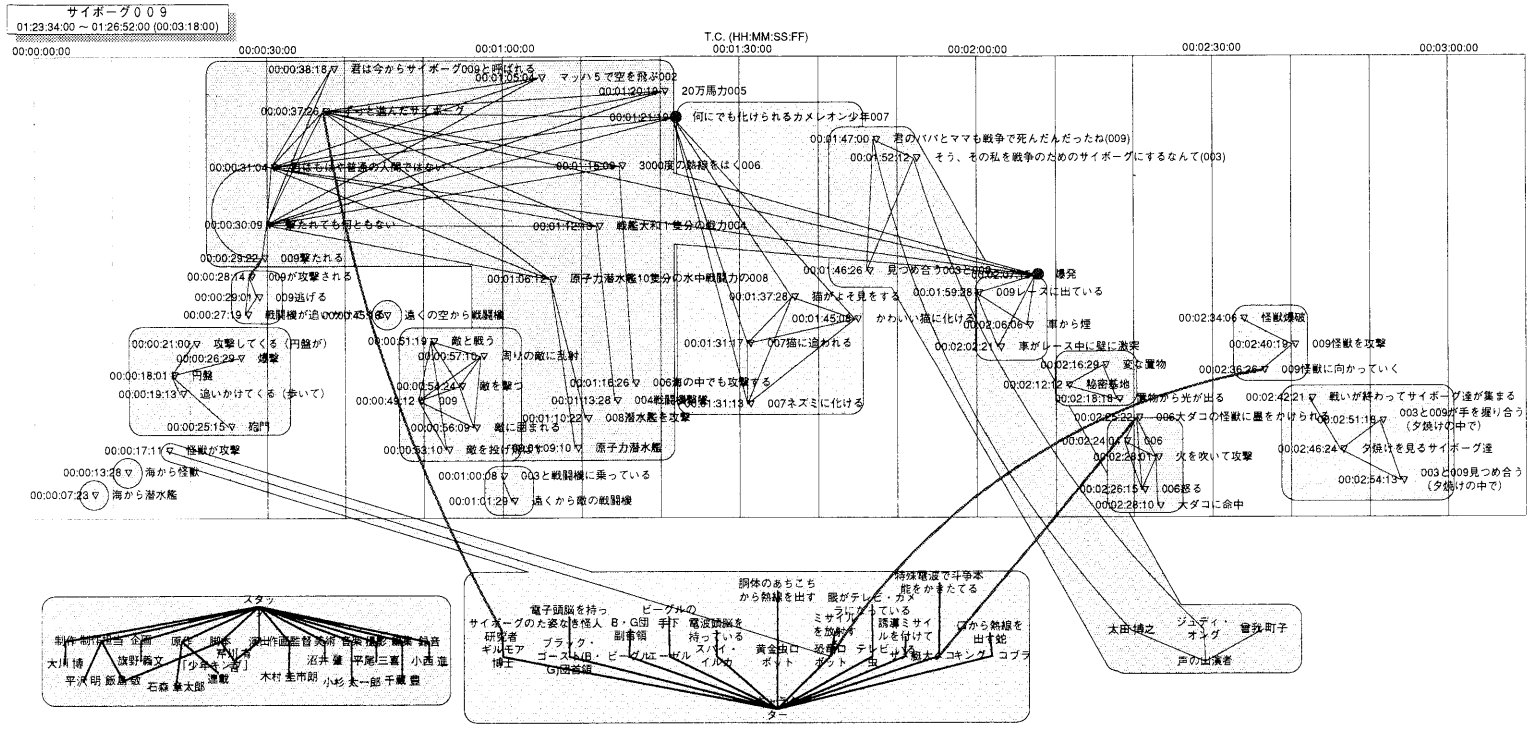


図２： 時刻印付ノードリンクグラフによる記述および検索



● カット点

— 橋

— 分断できない関連
（橋にできない辺）

＊ クラスタリングの際、常識リンクおよび汎化リンクは含まない。

図3： 時刻印付ノードリンクのクラスター化