

並行編集環境における XML文書の一貫性維持方式



鳥井 修



木村 哲郎



瀬川 淳一

(株) 東芝 研究開発センター

雑感

- 皆さんのシナップスを興奮させられるかどうか心配なところではありますが (DEWS, PYIWIT とともに最終セッション!), これも交流の一環とらえ
ご指導ご鞭撻のほどよろしくお願いします
-
- バックグラウンドが様々な人達の話の聞くことは大いに刺激になる、でも一人で飯を食うのはちょっと寂しい
-
- 学会というところは、相変わらず大学関係者と NTT の集まりのようだなあ
- dbjapan のメーリングリストへの企業からの登録が減っている (横田先生)、企業からの情報のディスクロージャーが米国に比べて圧倒的に少ない (喜連川先生) 済みません
-
- やぱり研究に対する真摯な姿には感銘を受けるものがある、いいなと思った研究が多数あったし、パネルディスカッションでは年甲斐もなく感動してしまった
- 一方、10 分のところ一時間喋ってしまってどうして健全な『コミュニティ』が維持できるのか分からない
-
- NOVA に行かなければ
- VLDB 行こうかな?
-
- 教授の仕事の『下請け』をしたことは、ただの一度もなかったけどなあ

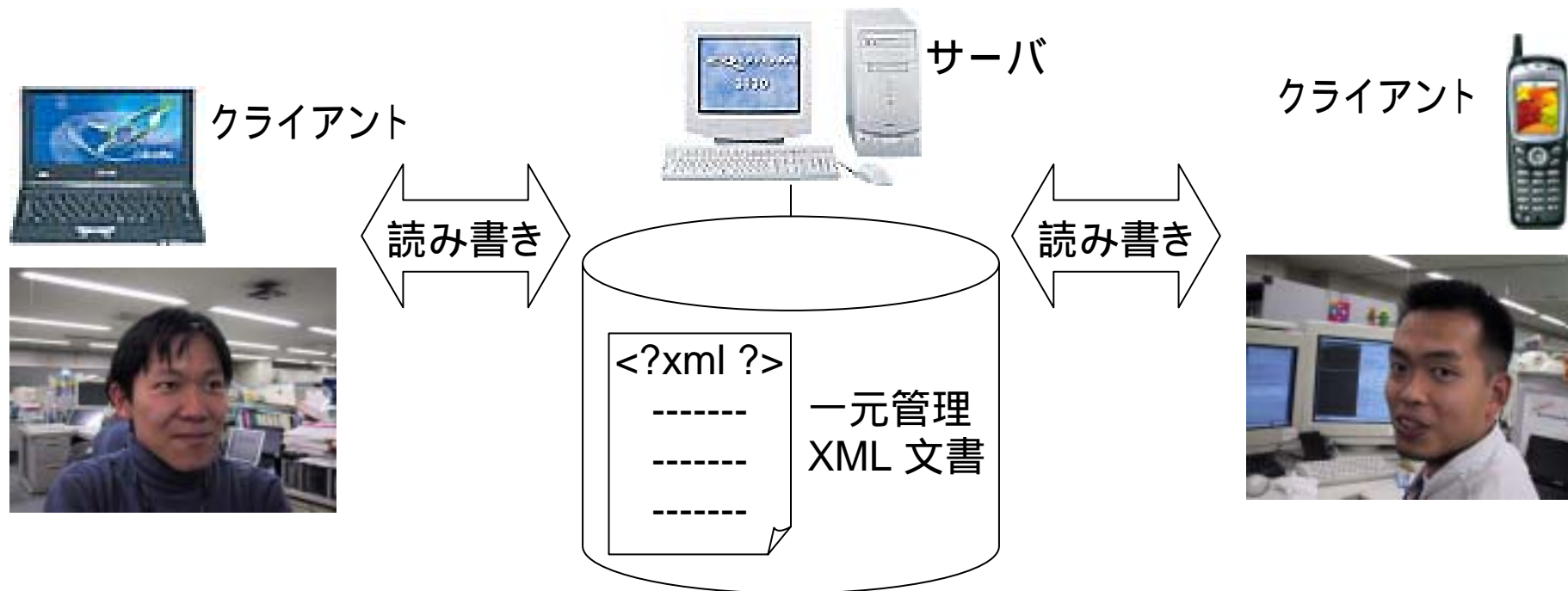
発表の概要

- 発表 (15 分)
 - 背景 (2 分)
 - XML文書の並行編集システム (2 分)
 - XML文書の構造情報編集の分割 (10 分)
 - 結論 (1 分)
- 質疑応答 + おまけ (15 分)

発表の概要

- 発表 (15 分)
 - 背景 (2 分)
 - XML文書の並行編集システム (2 分)
 - XML文書の構造情報編集の分割 (10 分)
 - 結論 (1 分)
- 質疑応答 + おまけ (15 分)

やりたいこと



同一の XML 文書を
皆で編集するシステムを作りたい

難しいところ

XML 文書は単なる文書と異なり、
構造情報を持つ

- そして、サーバ上の XML 文書は、
- (1) (構造情報が) 常に well-formed でなければいけない
 - (2) (構造情報が) 編集者の編集意図に近いものでなければいけない

研究の主眼

編集者がクライアント上で行った編集を、
『適切な』編集の単位に分割し、この単位ごとに、
(1) コンフリクト判定、
(2) サーバへの反映、
を行えばよい

しかし、実際には『**構造情報**』に関する編集を、
『適切な』単位に分割するのは難しい

『適切な単位って何だろう？』と
考えたのが本研究

発表の概要

- 発表 (15 分)
 - 背景 (2 分)
 - XML文書の並行編集システム (2 分)
 - XML文書の構造情報編集の分割 (10 分)
 - 結論 (1 分)
- 質疑応答 + おまけ (15 分)

文字情報と構造情報の分離

```
<?xml version="1.0" ?>
```

```
<論文>
```

```
<節>
```

```
<段落>
```

```
  並行編集システムとは ...
```

```
</>
```

```
</>
```

```
<節>
```

```
<段落>
```

```
  XT グラフとは ...
```

```
</>
```

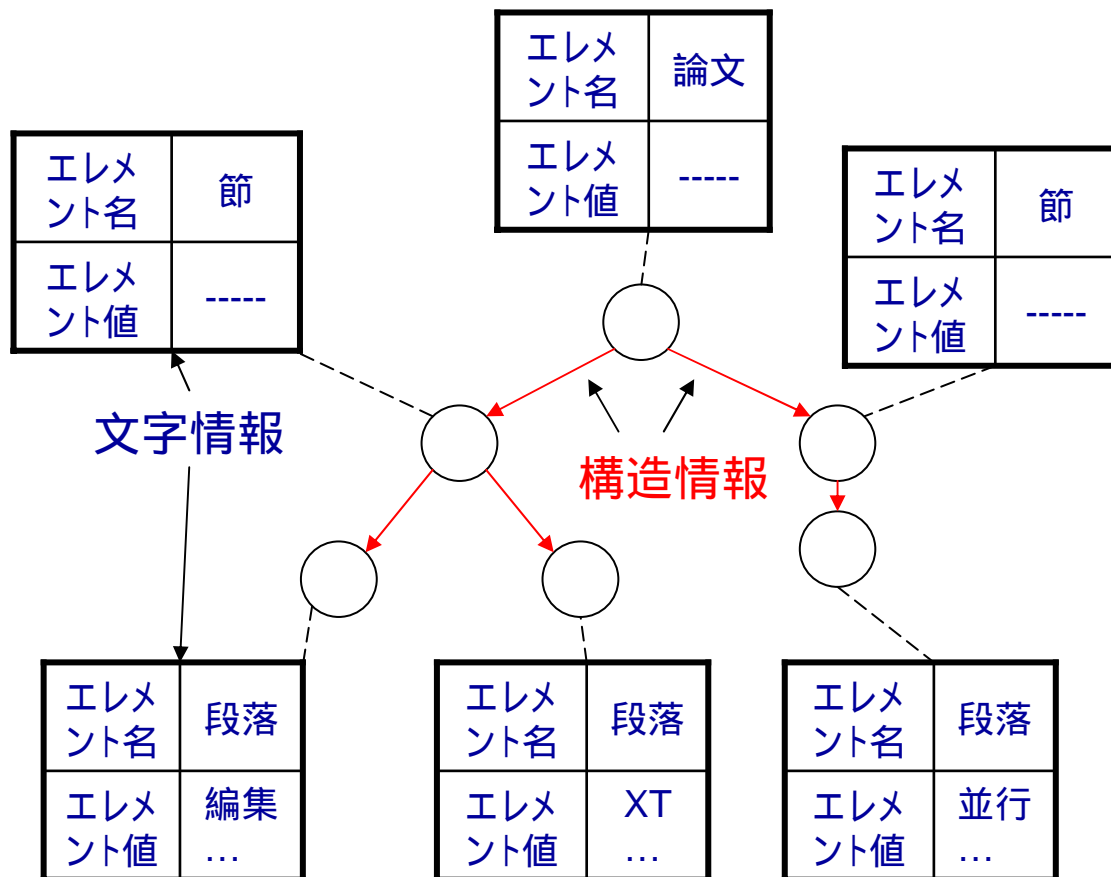
```
<段落>
```

```
  編集グラフとは ...
```

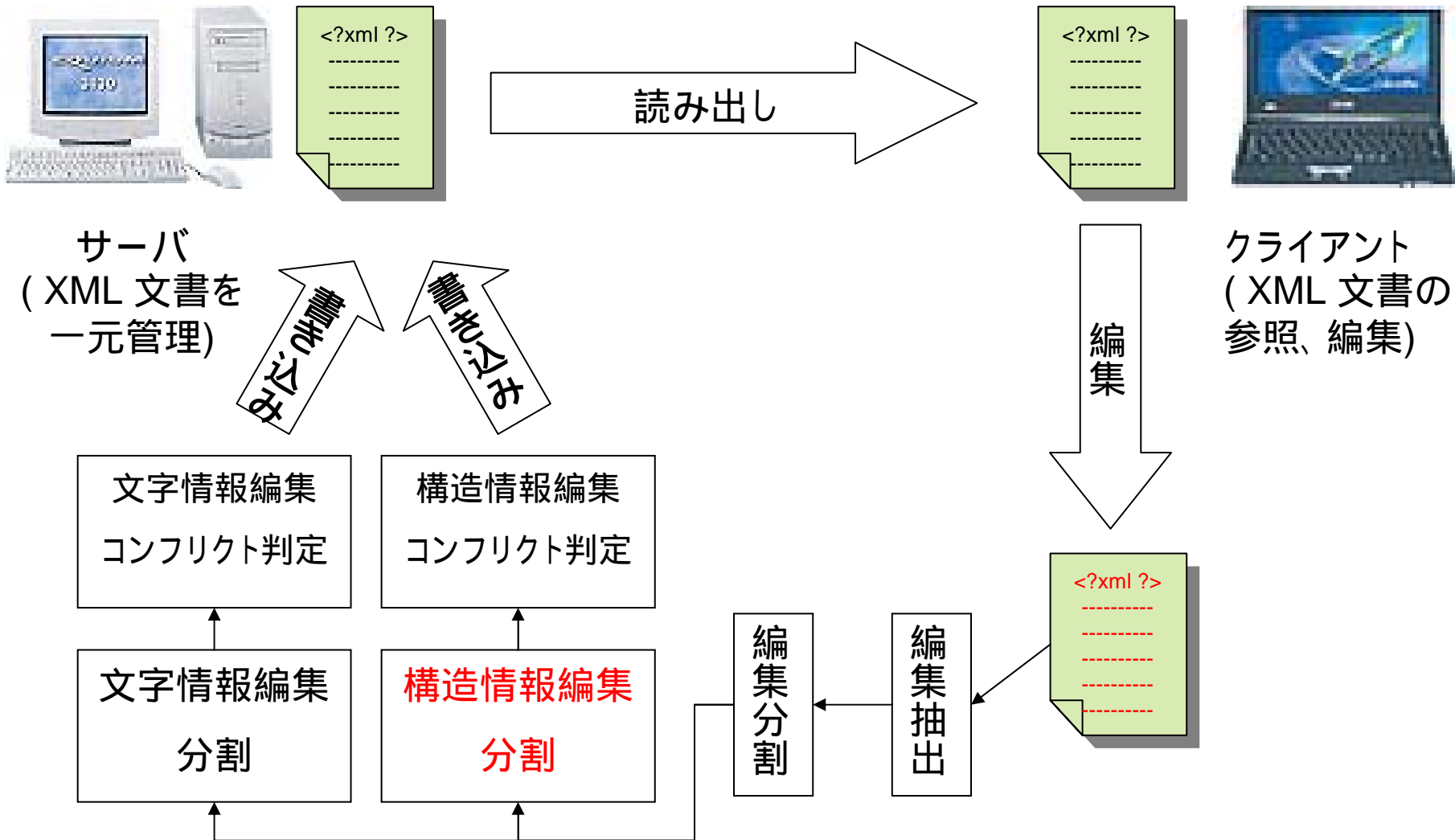
```
</>
```

```
</>
```

```
</>
```



システムの動作



XML文書の並行編集システム

発表の概要

- 発表 (15 分)
 - 背景 (2 分)
 - XML文書の並行編集システム (2 分)
 - XML文書の構造情報編集の分割 (10 分)
 - 結論 (1 分)
- 質疑応答 + おまけ (15 分)

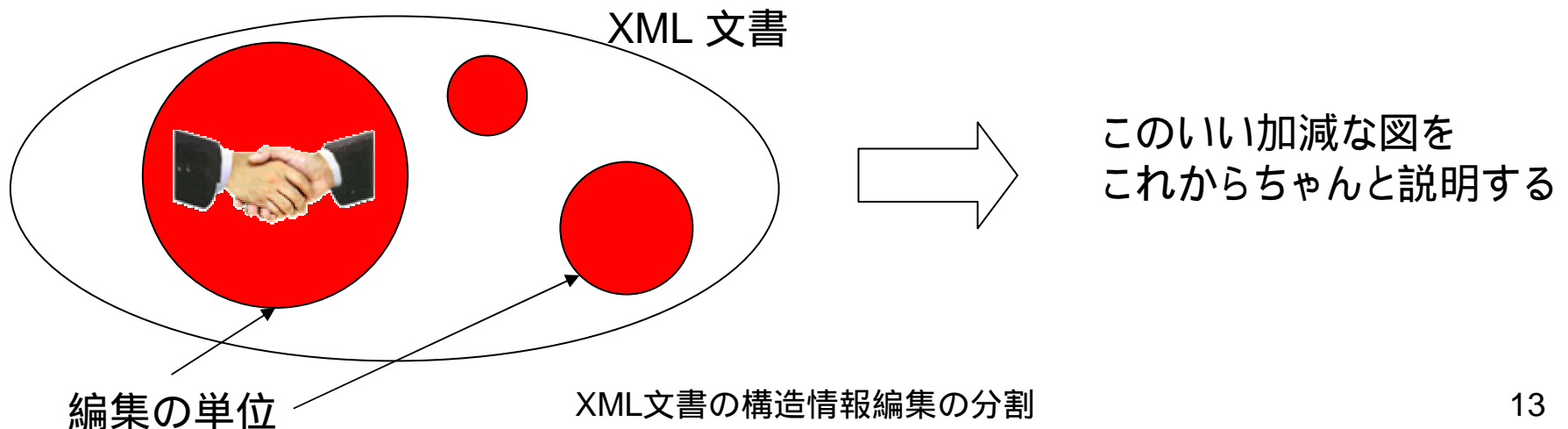
やりたいこと

XML 文書の構造情報に関する編集を、
『適切な』編集の単位に分割すること

『適切な』編集の単位って何？

編集の単位が『適切である』とは
以下の3つの条件を満たしていること

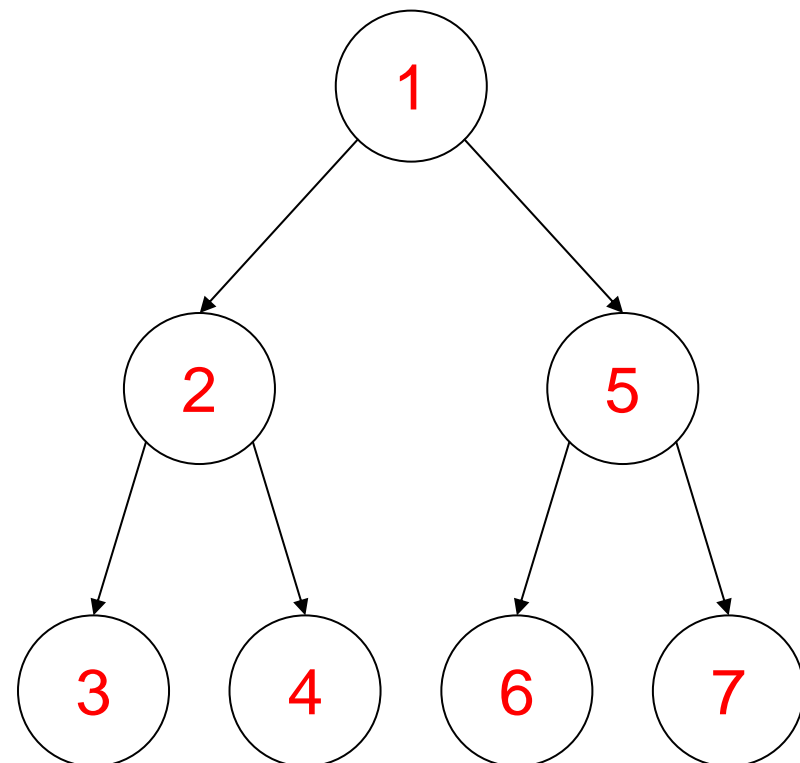
- 『整合性』: XML 文書の well-formedness を崩さない
- 『独立性』: 編集の単位間の関連がない
- 『極小性』: これ以上小さくできない



XML 文書の構造情報の表現 (従来)

```
<?xml version="1.0" ?>
<論文 id="1">
  <節 id="2">
    <段落 id="3">システムの概要</>
    <段落 id="4">システムの詳細</>
  </節>
  <節 id="5">
    <段落 id="6">分割の概要</>
    <段落 id="7">分割の詳細</>
  </節>
</論文>
```

id の存在を前提



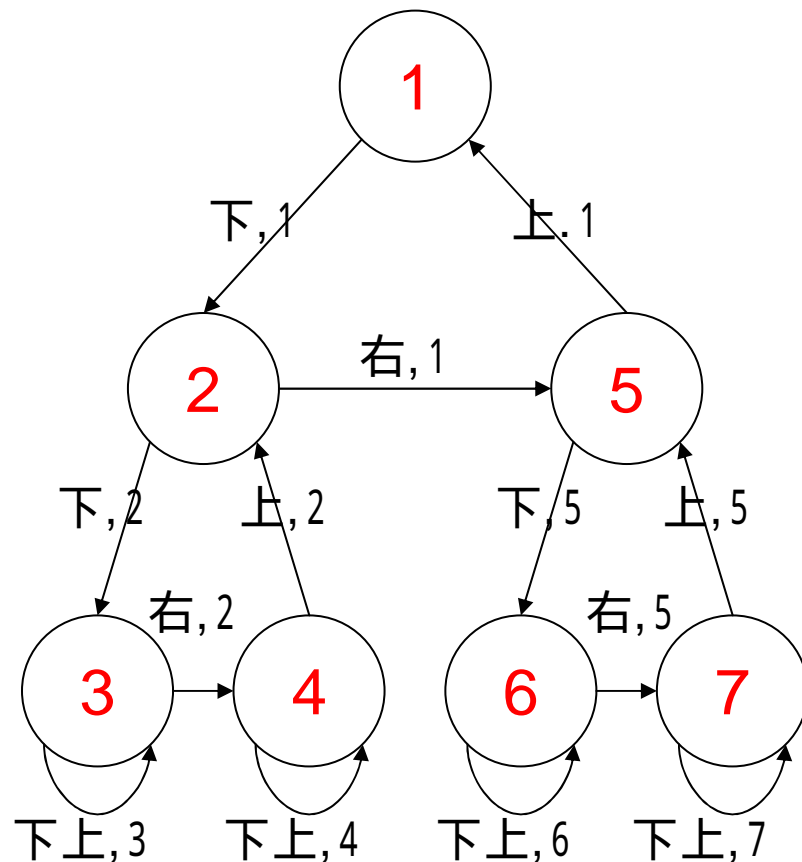
XML 文書の構造情報の表現 (XT グラフ)

XT グラフ = XML tree グラフ or XML topology グラフ

```

< ?xml version="1.0" ? >
< 論文 id="1" >
  < 節 id="2" >
    < 段落 id="3" > システムの概要 < / >
    < 段落 id="4" > システムの詳細 < / >
  < /節 >
  < 節 id="5" >
    < 段落 id="6" > 分割の概要 < / >
    < 段落 id="7" > 分割の詳細 < / >
  < /節 >
< /論文 >

```



『編集』と『編集グラフ』の関係

編集グラフの黒サブグラフ
= 編集前の XML 文書の XT グラフ

編集グラフのすべての辺を
スイッチした結果得られる
編集グラフの黒サブグラフ
= 編集後の XML 文書の XT グラフ

編集
= **編集グラフの辺のスイッチ**

編集グラフの黒サブグラフ
= 編集中の XML 文書

やりたかったこと

XML 文書の構造情報に関する編集を、
『適切な』編集の単位に分割すること

これを今まで導入してきた、
データ構造を用いて言い直すと...

『適切な』編集グラフ分割

XML 文書の構造情報に関する編集を、
『適切な』編集の単位に分割すること

=

編集グラフ中のすべての辺を、
『適切な』辺グループに分割すること

=

編集グラフ中のすべての辺を、
『well-formedness を維持し』、
『互いに関連がうすく』、
『これ以上細かく分割することが困難な』
辺グループに分割すること

=

『適切な』編集グラフ分割

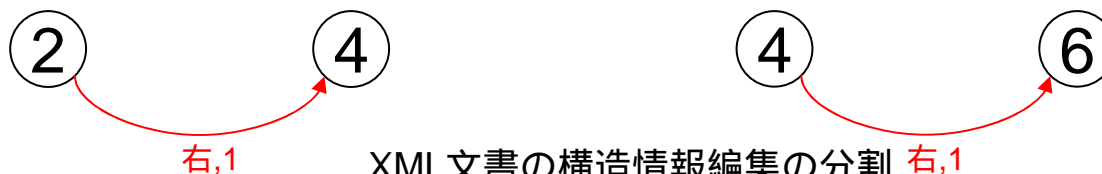
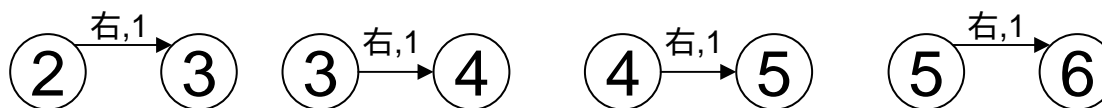
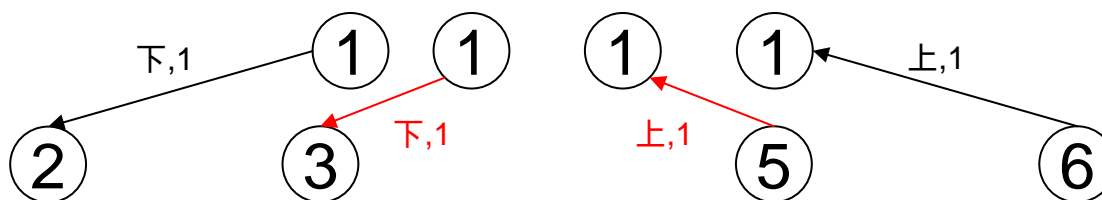
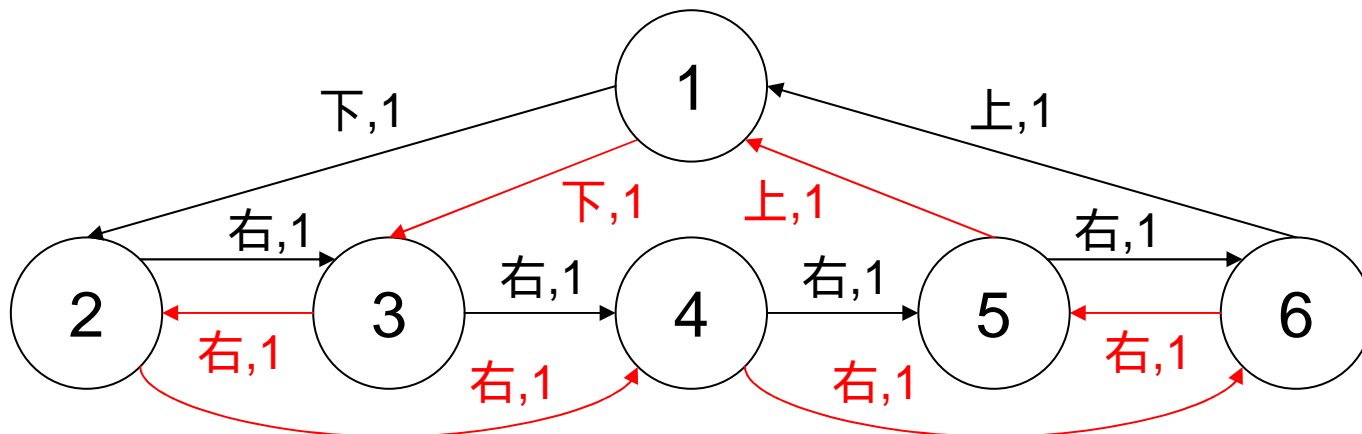
適切な編集グラフ分割を求める アルゴリズム

1. 編集グラフの作成
2. 編集グラフを『最小の』編集の単位に分割
3. 『最小の』編集の単位をまとめ、『適切な』編集の単位を作成する

適切な編集グラフ分割を求める アルゴリズム

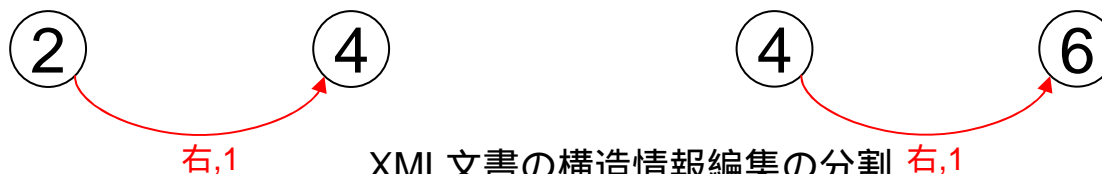
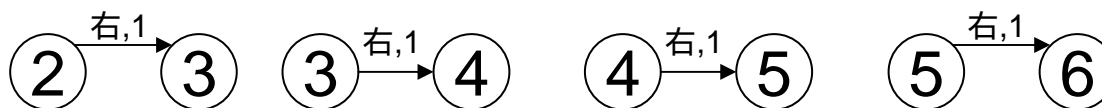
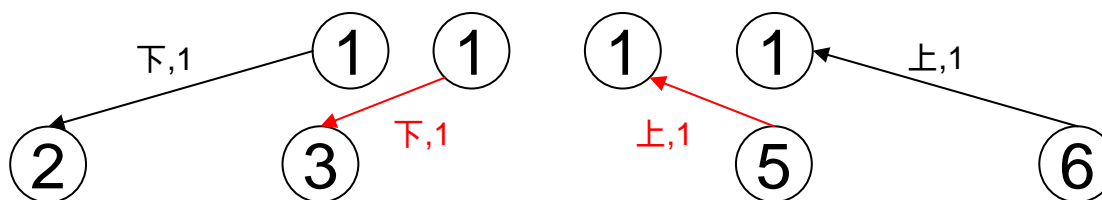
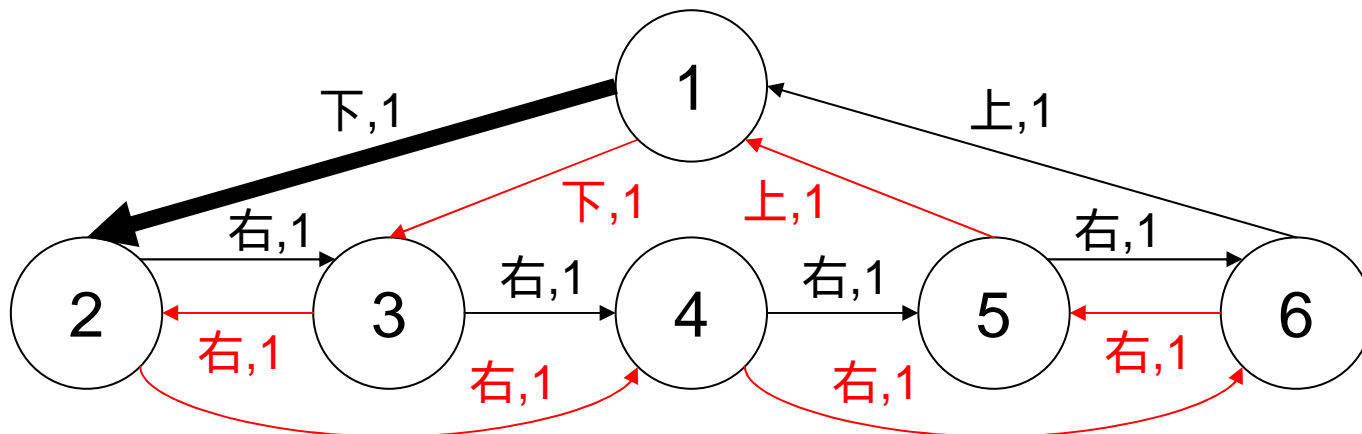
1. 編集グラフの作成
2. 編集グラフを『最小の』編集の単位に分割
3. 『最小の』編集の単位をまとめ、『適切な』編集の単位を作成する

(第一の) まとめあげ



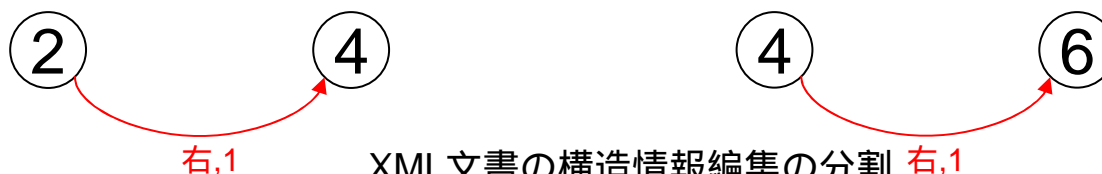
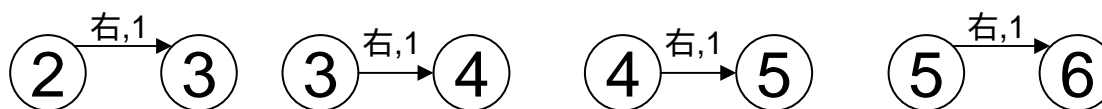
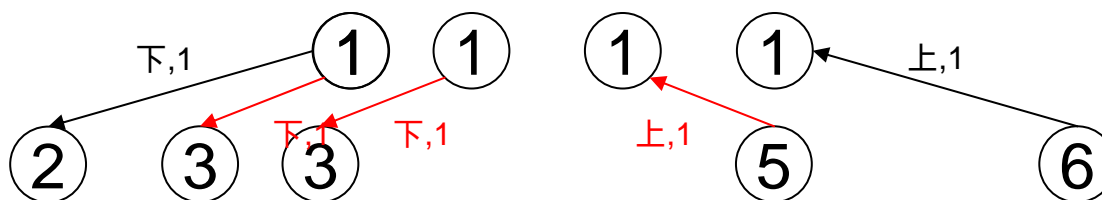
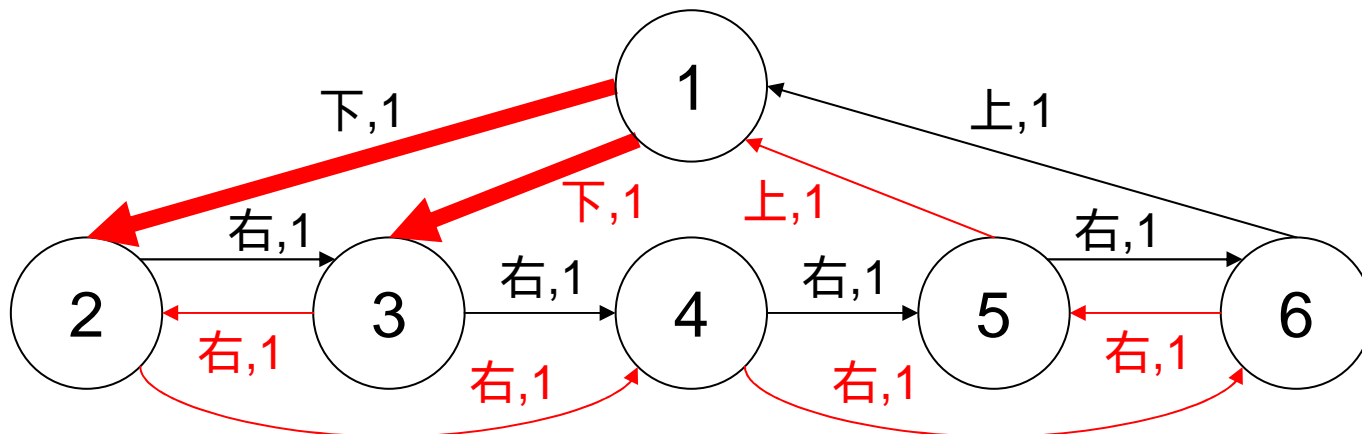
XML文書の構造情報編集の分割

(第一の) まとめあげ



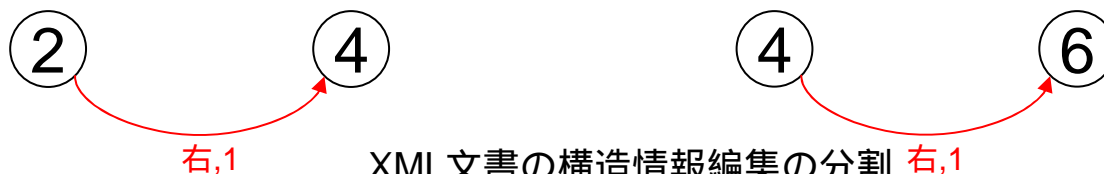
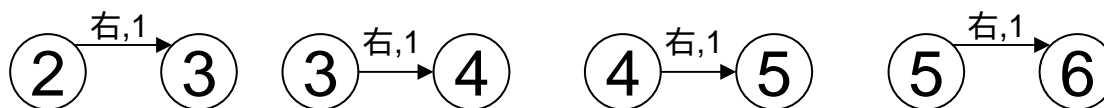
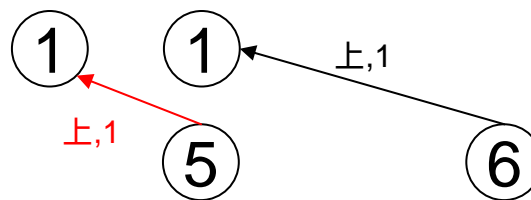
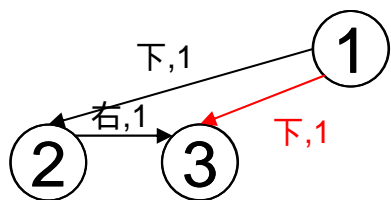
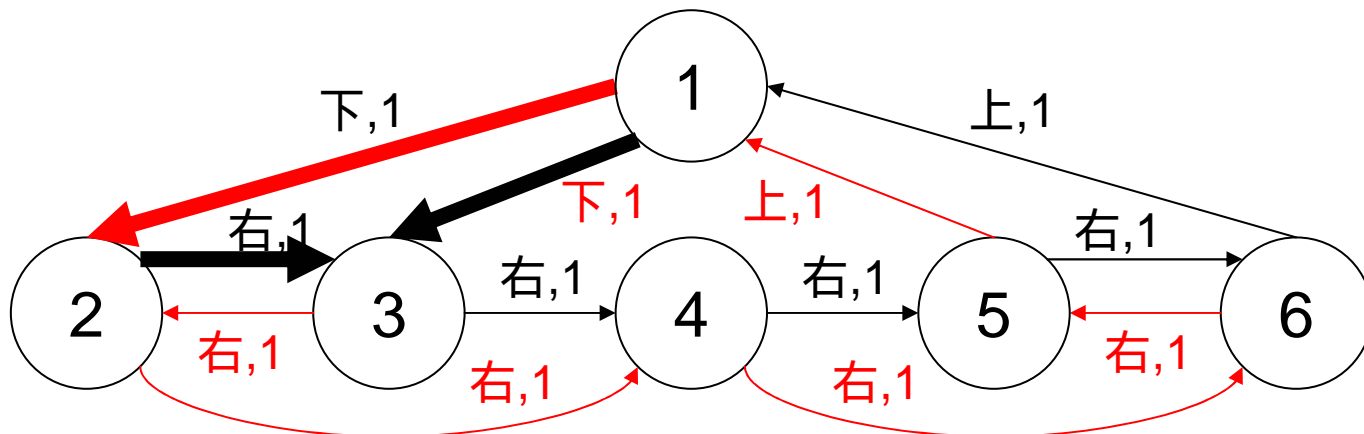
XML文書の構造情報編集の分割

(第一の) まとめあげ



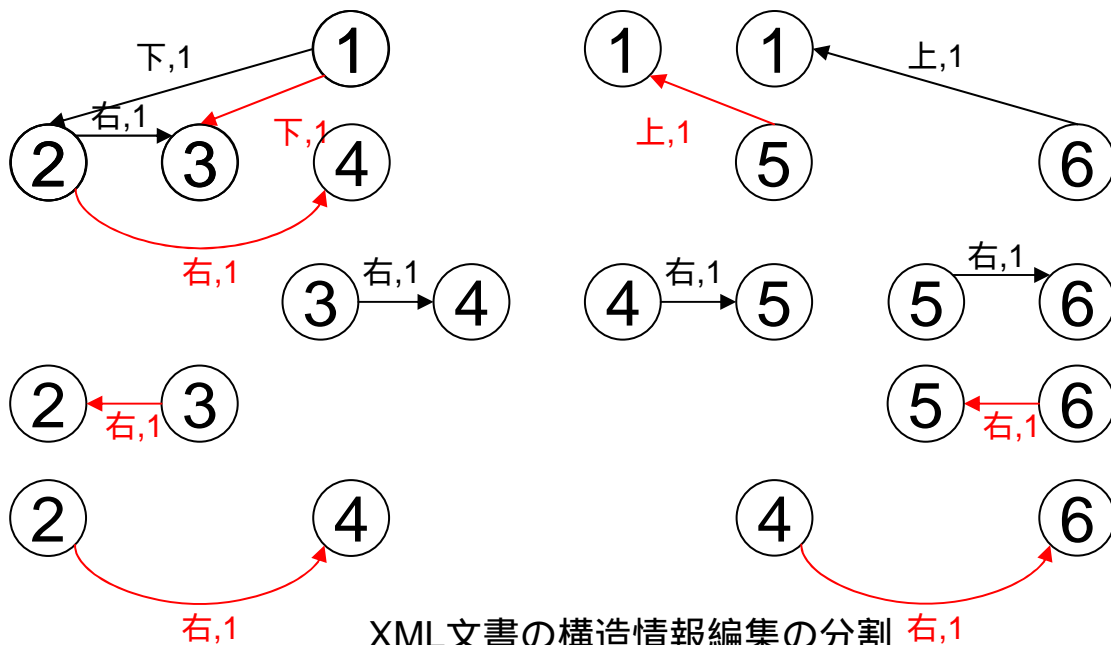
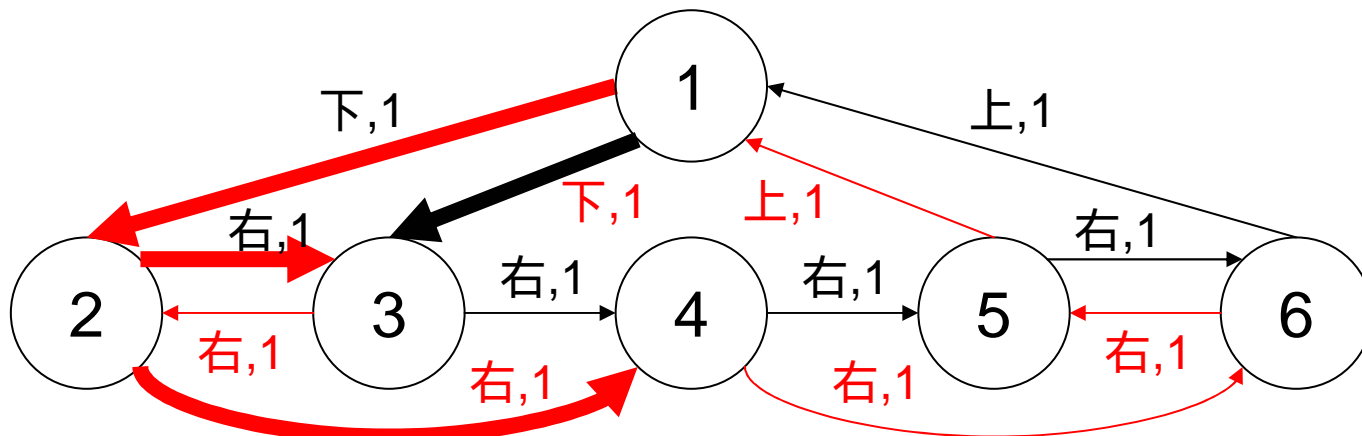
XML文書の構造情報編集の分割

(第一の) まとめあげ



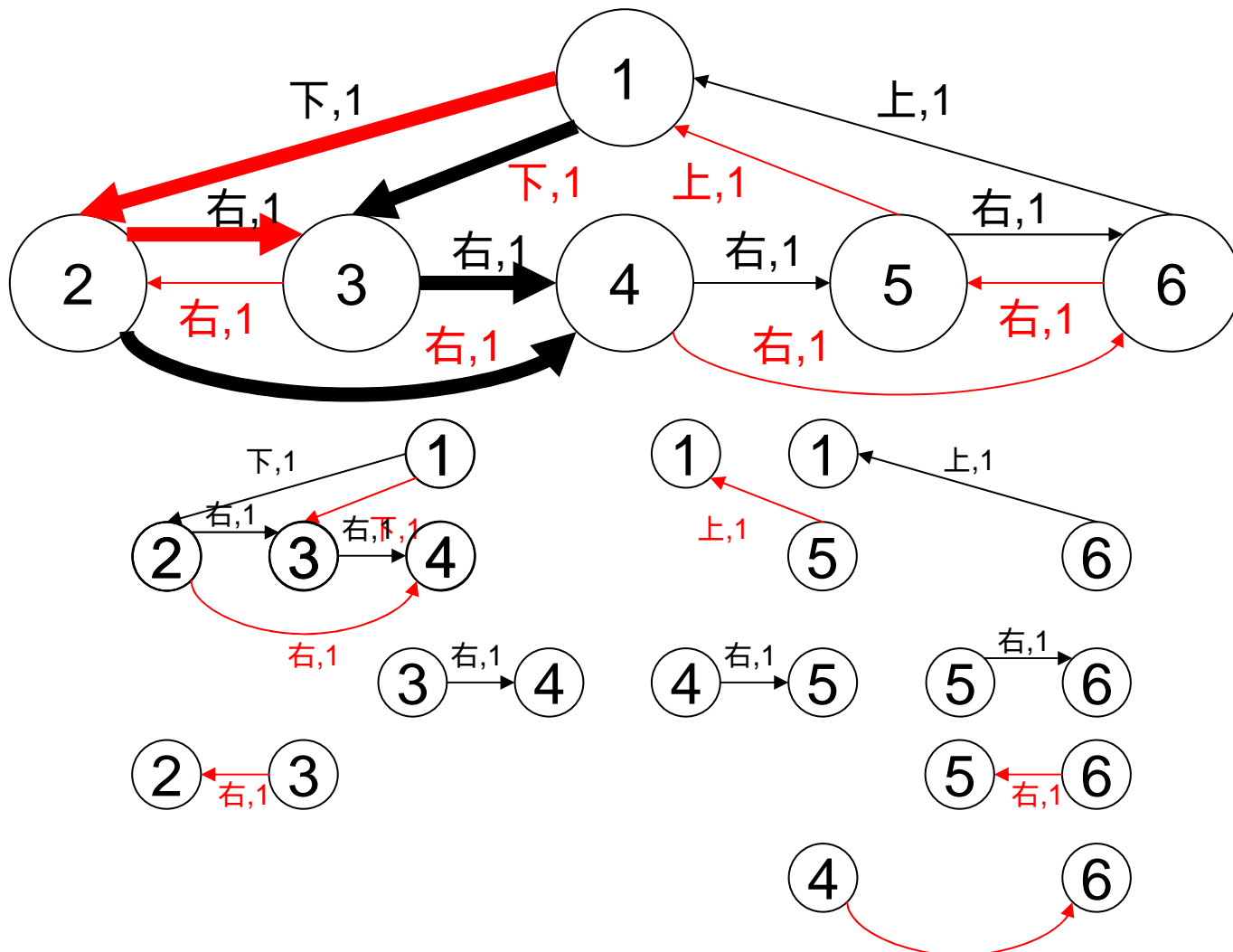
XML文書の構造情報編集の分割

(第一の) まとめあげ

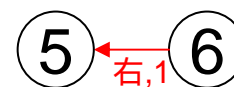
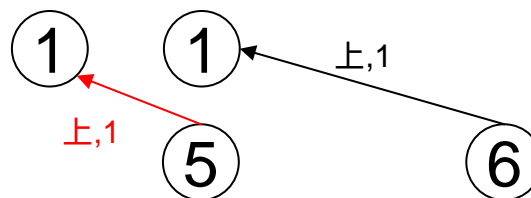
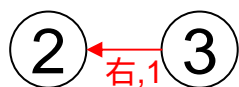
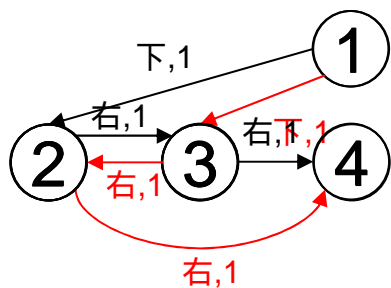
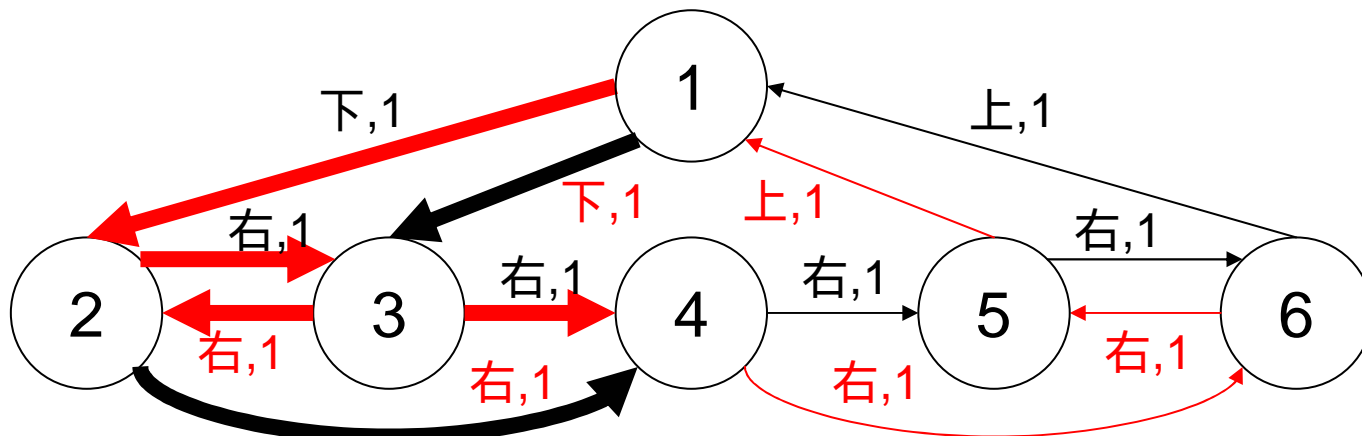


XML文書の構造情報編集の分割 右,1

(第一の) まとめあげ

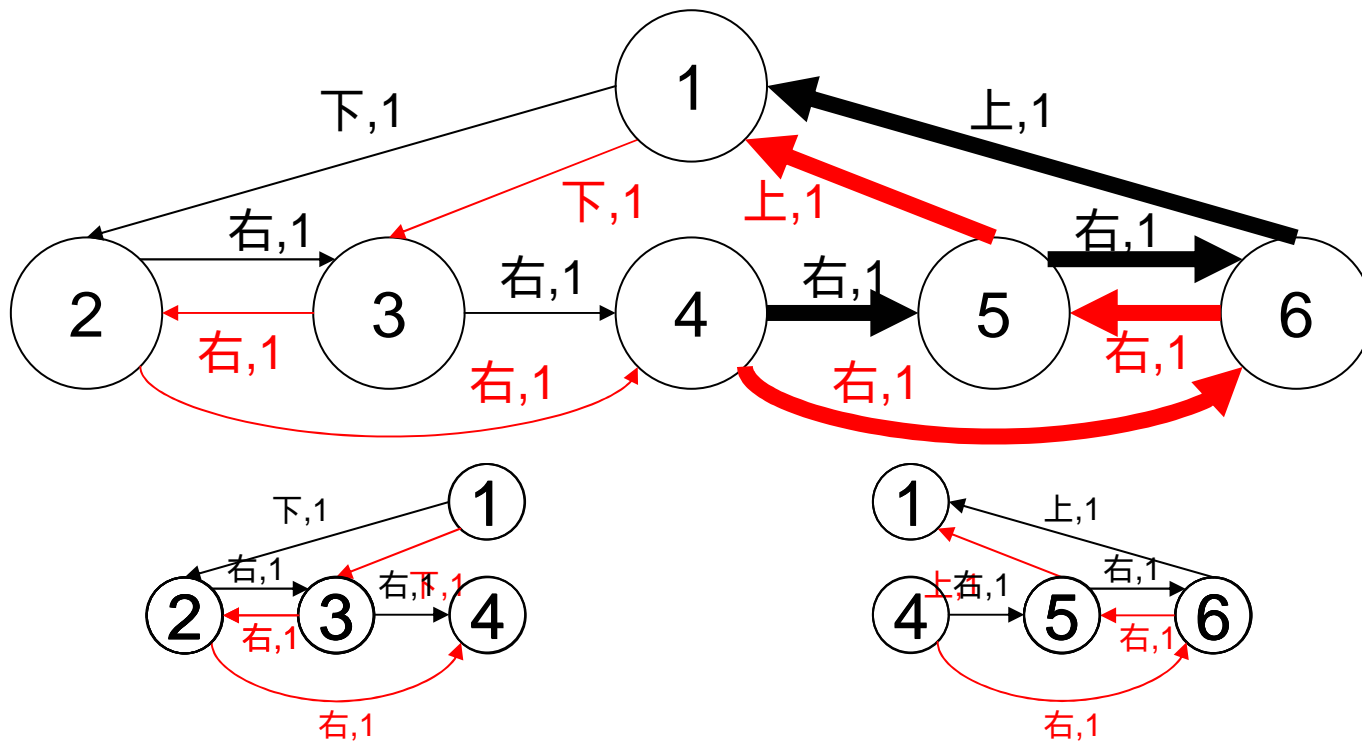


(第一の) まとめあげ



XML文書の構造情報編集の分割 右,1

(第一の) まとめあげ



交互輪 = オルターネイティングリング

まとめあげ (続き)

本当は第二、第三、第四の
まとめあげと続くが本日は省略

こんな調子で、
『適切な』編集グラフ分割、
つまりXML 文書の構造情報に関する編集の
『適切な』編集の単位への分割が
得られる

発表の概要

- 発表 (15 分)
 - 背景 (2 分)
 - XML文書の並行編集システム (2 分)
 - XML文書の構造情報編集の分割 (10 分)
 - 結論 (1 分)
- 質疑応答 + おまけ (15 分)

結論

- サーバ上で一元管理される XML 文書を編集者の意図に近い形式に維持するための一貫性維持方式を提案した
- XML 文書の構造情報に関する編集を、適切な単位に分割する手法を確立した
- 本方式を用いたシステムは現在実装中、実証実験を通じて問題点の洗い出しを行う予定

発表の概要

- 発表 (15 分)
 - 背景 (2 分)
 - XML文書の並行編集システム (2 分)
 - XML文書の構造情報編集の分割 (10 分)
 - 結論 (1 分)
- 質疑応答 + おまけ (15 分)

質疑応答 + おまけ

- ご指導、ご鞭撻よろしくお願いします
- 座長の森嶋先生は必ずやアグレッシブな質問を下さるはず
- しかし、『質問がないようなので座長から...』というのは寂しい...
- 恥ずかしながら、査読結果で頂いた『論文の読み易さ』は『7点』と『6点』... 絶対に今質問頂いた方がよい

どうして XML 文書？

『XML 文書』を前面に出しているのは、
単に XML 文書がはやりであるから

『並行編集環境における
ツリーの一貫性維持方式』
というタイトルでも別に構わない、
と言うかその方が正確

関連研究

- (1) CVS
- (2) マルチユーザを前提とした、XML 文書のバージョンングシステム『そのものずばり』は見付けることができず
求む情報

XML 文書の差分

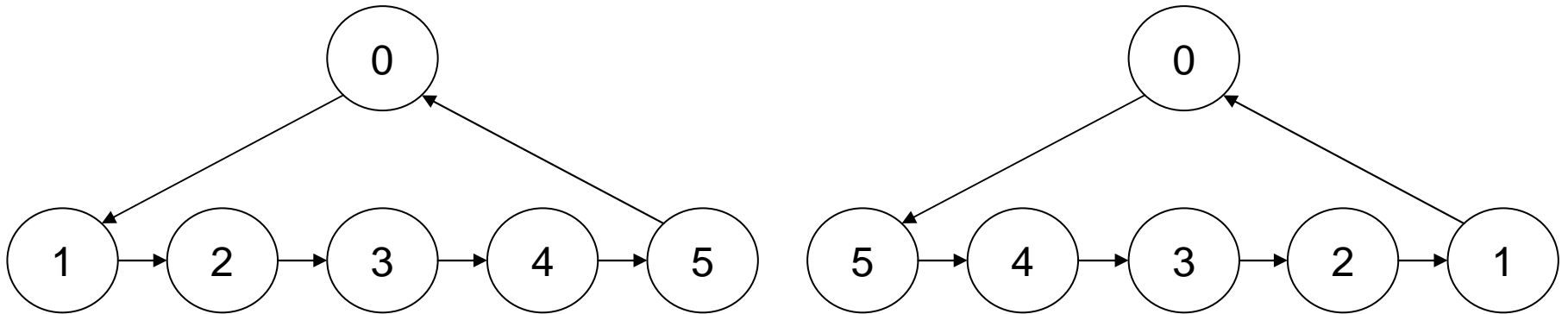
天笠先生がいらっしゃったら、『edit distance との関係は?』というご質問をなさるはず

- XML 文書の差分発見の手法に関しては、結構多くの研究が存在する
- その多くが、二つの XML 文書中の対応するエレメントを発見 (し、差分を『作成』、『削除』、『移動』編集でシミュレート) する方法
- しょせんは、エレメントに id をふるためだけの手法であり、コンカレント・バージョンング・システムを構築するためにはもう一ひねり必要

XML 文書の差分 (関連研究リスト)

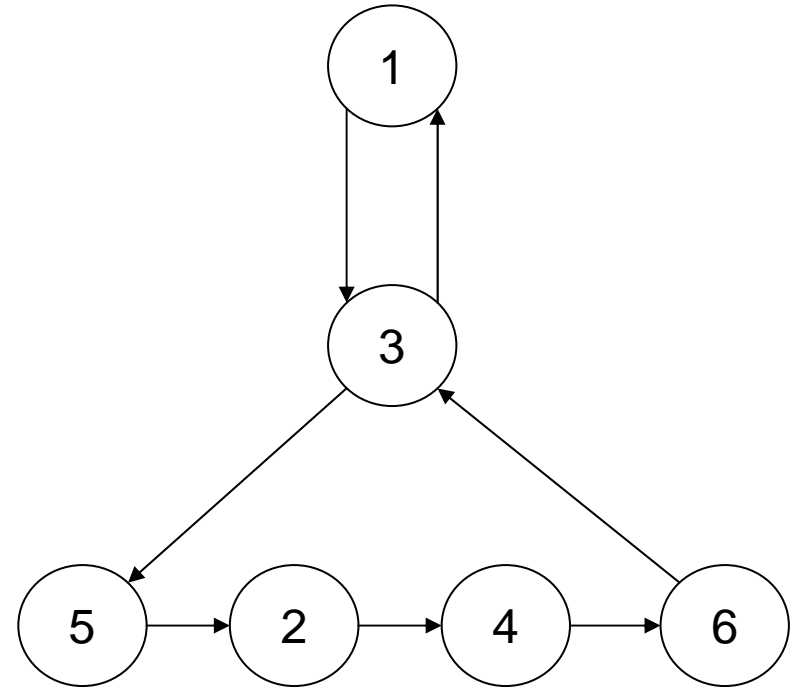
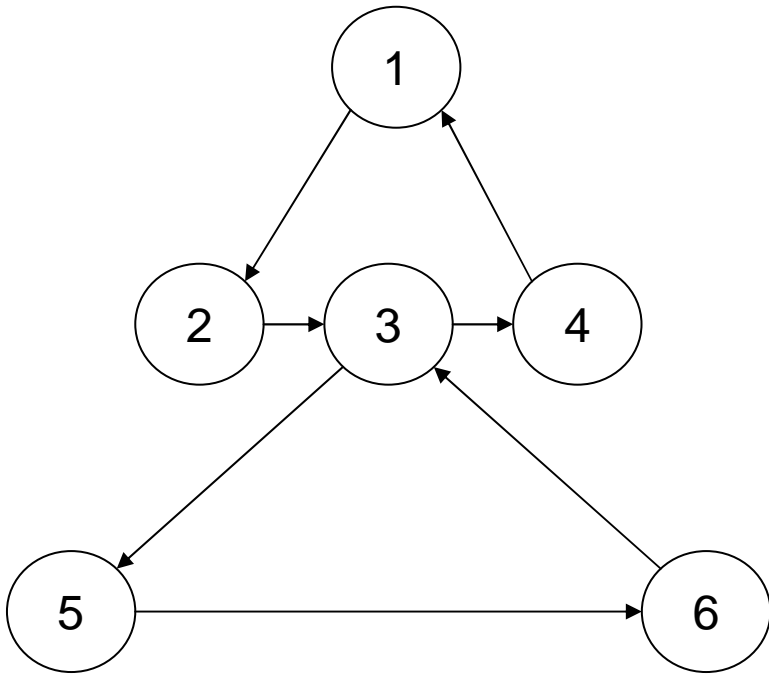
- A1-5: 児島先生ほか、Colored Bottom-up DataGuide による半構造データのための差異発見・可視化機構
 - Seung-Jin Lim, Yiu-Kai Ng, An Automated Change-Detection Algorithm for HTML Documents Based on Semantic Hierarchies
 - Roy Goldman, Jennifer Widom, DataGuides: Enabling Query Formulation and Optimization in Semistructured Databases
- C2-3: 鈴木先生ほか、構造の異なる文書データを同じ構造に変換するアルゴリズムについて
 - Chawathe, A. Rajaraman, H. Gracia-Molina, and J. Widom, Change Detection in Hierarchically Structured Information
 - K. Zhang and D. Shasha, Simple Fast Algorithms for the Editing Distance Between Trees and Related Problems
- A4-3: 宮崎先生ほか、XMLデータベースからの動的Web ページ生成環境における変更検出・通知方式
 - Seung Jin Lim ほか、An Automated Change Detection
- A5-3: 横山先生ほか、XML パーサを考慮した応用向き関係データベース構成法

XML 文書の差分の違和感 (1)



動いたのはいったい誰 ?

XML 文書の差分の違和感 (2)

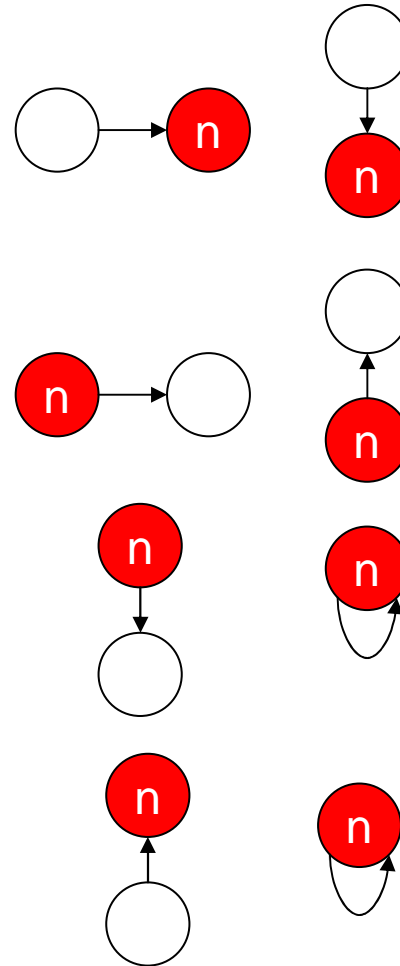


どうしてまとめあげは 四種類あるの？

あるグラフが
XT グラフである条件が
四種類あるから

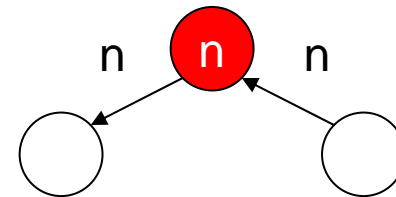
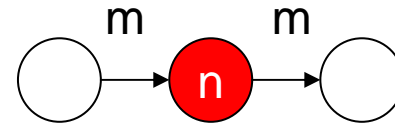
条件一

1. (r 以外の) 任意の頂点 n を終了頂点とする右向きまたは下向き辺が、合計一つ存在する。
2. (r 以外の) 任意の頂点 n を開始頂点とする右向きまたは上向き辺が、合計一つ存在する。
3. 任意の頂点 n を開始頂点とする下向きまたは下上向き辺が、合計一つ存在する。
4. 任意の頂点 n を終了頂点とする上向きまたは下上向き辺が、合計一つ存在する。



条件二

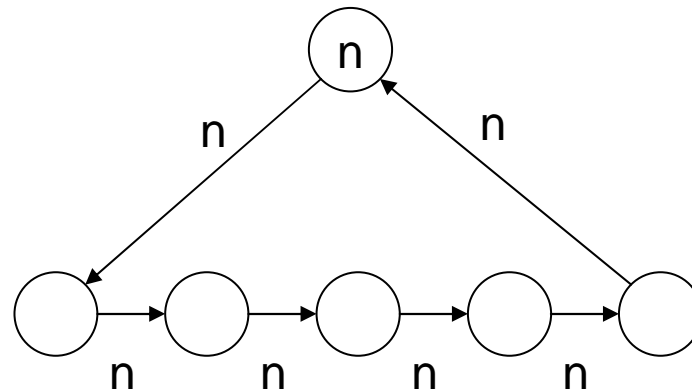
1. (r 以外の) 任意の頂点 n の左辺と右辺のラベルは同一である
2. 任意の頂点 n の左下辺と右下辺のラベルはともに n である



条件三

水平ループは、
このループ中の辺ラベルと
同一の頂点 id を保持する頂点を
必ず通過する

水平ループ
= 同一のラベルを保持する辺のみから構成されるループ



条件四

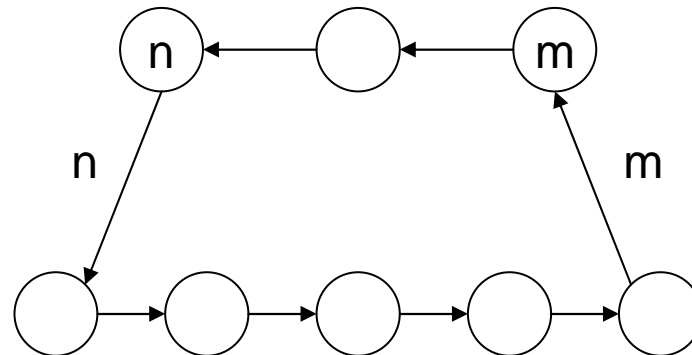
プリミティブな垂直ループは、存在しない

プリミティブなループ

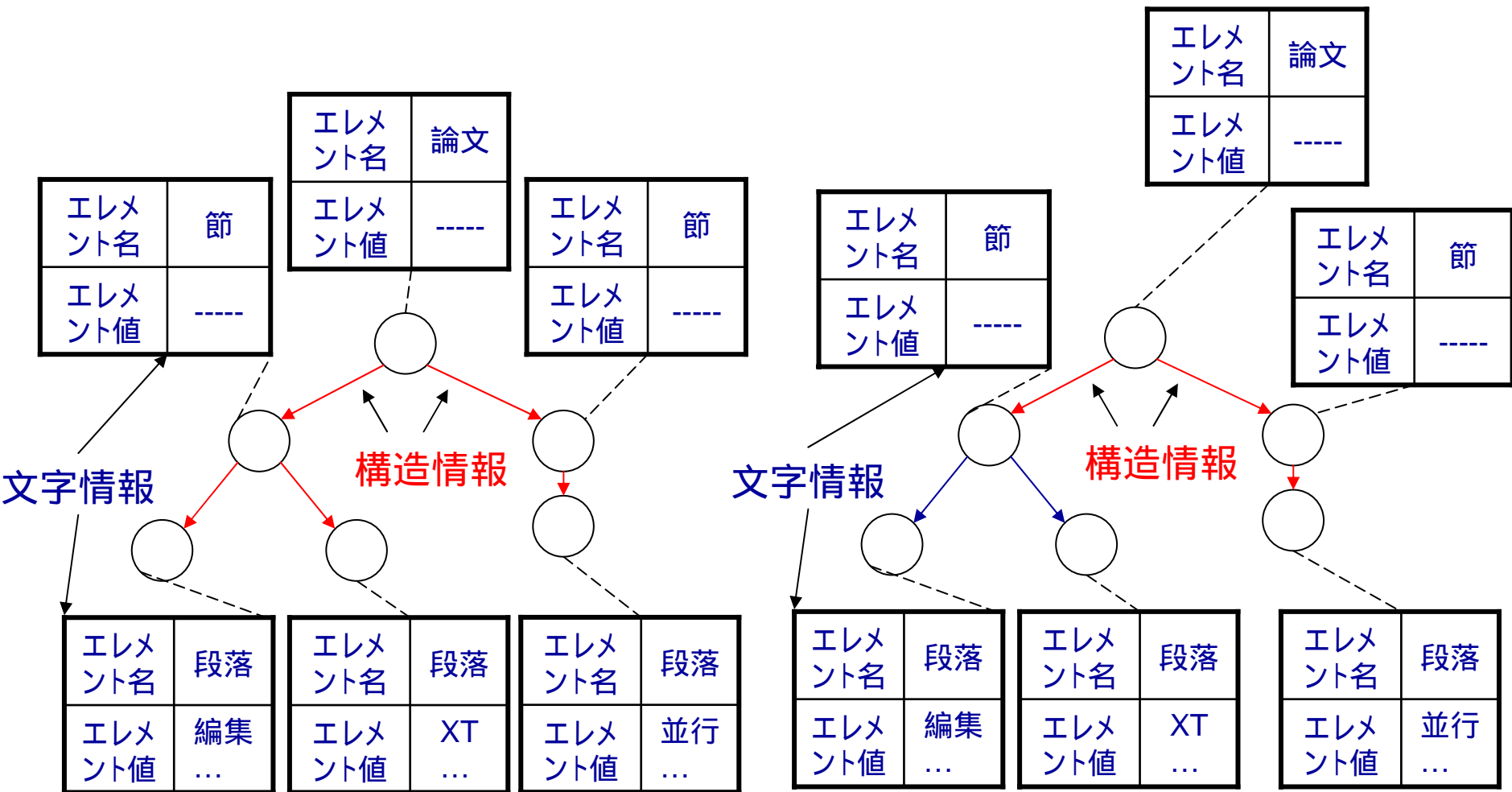
= 同一のノードを一回のみ通過するループ

水平ループ

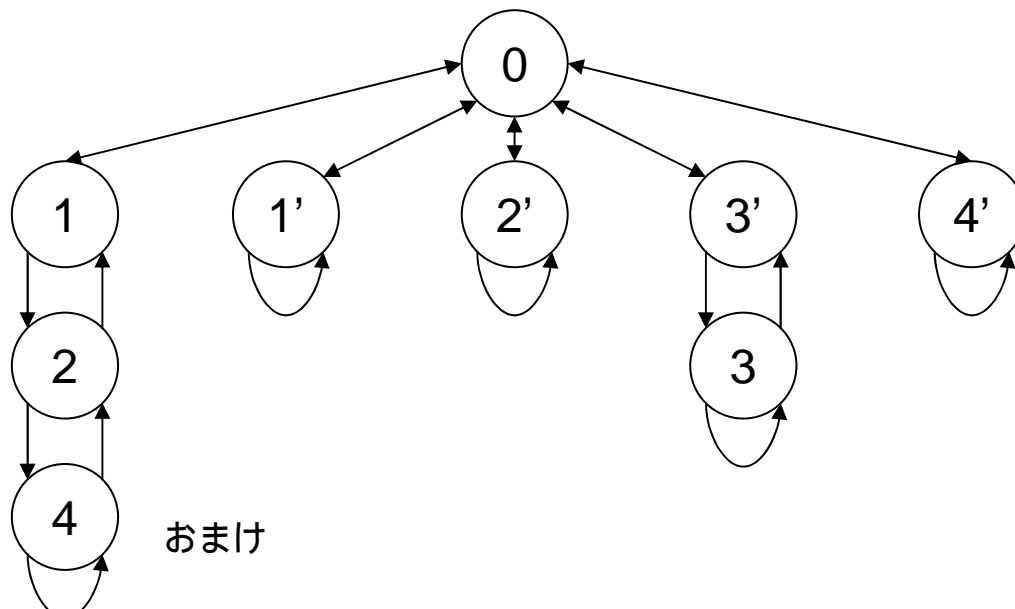
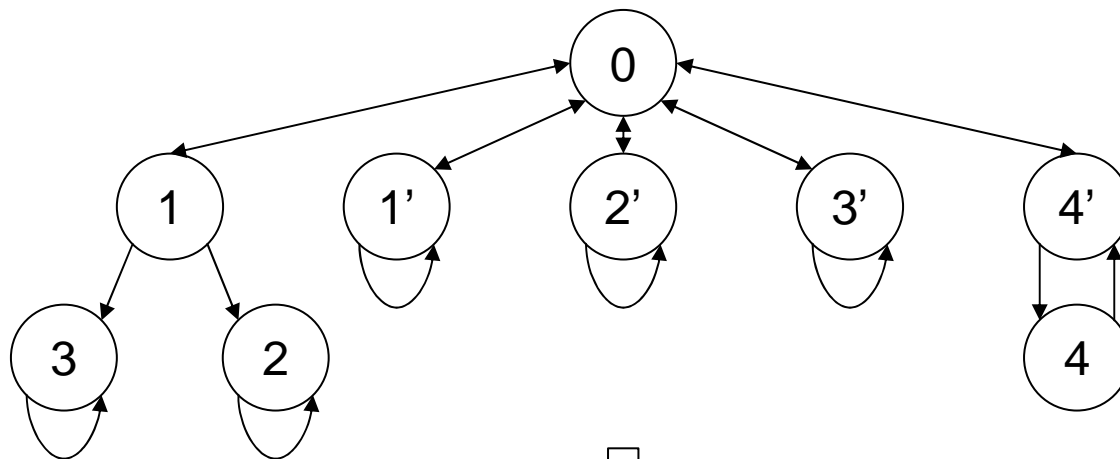
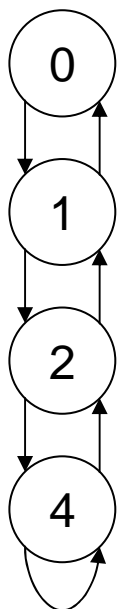
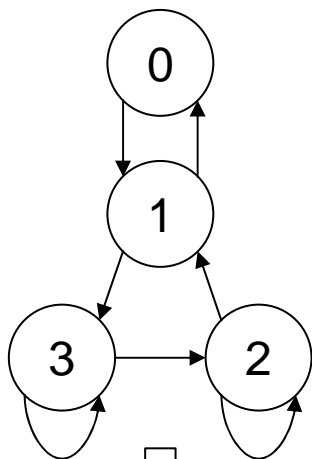
= 同一のラベルを保持する辺のみから構成されるループ



文字列と構造情報は常に独立？



XT グラフの拡張



おまけ