SCIS 2002 The 2002 Symposium on Cryptography and Information Security
Shirahama, Japan, Jan. 29 - Feb.1, 2002
The Institute of Electronics, Information and Communication Engineers

XML におけるステガノグラフィ手法の提案
A Proposal on Steganography Methods using XML

井上 信吾[*1] 村瀬 一郎[*1] 滝澤 修[*2] 松本 勉[*3] 中川 裕志[*4]
Shingo Inoue Ichiro Murase Osamu Takizawa Tsutomu Matsumoto Hiroshi Nakagawa

あらまし XML 文書に対するステガノグラフィ手法を提案する。XML が Web ページなどの多様な文書を記述可能な構造化テキストであることに注目し、テキスト中のマークアップ部と文書構造の定義部を対象に秘匿情報の埋め込みを行う方法を複数示す。手法は大きく 2 つのグループに分けられる。第 1 のグループに属する手法では XML 標準で許容されている記法のバリエーションを利用して秘匿情報の埋め込みを行う。第 2 のグループに属する手法は DTD 等のスキーマ記述の冗長性や、アプリケーションのコンテキスト上同義と見なされる記述を利用し、文書構造を改変して秘匿情報が埋め込まれた文書を作成するものである。

キーワード ステガノグラフィ、情報ハイディング、電子すかし、XML、テキスト、構造化文書

1 はじめに

現在、ネットワークを多様な形態・性質の情報が送受されるが、その中でもテキストは、多くのアプリケーションで一般的に送受されるデータ形式と言える。これら流通するテキストの多くは単なるプレインテキストではなくアプリケーション固有の書式やマークアップ言語で書かれた整形済データである。XML は情報のフォーマットを記述するメタ言語として知られているが、Web ページや商取引データなどの様々な文書の記述や、ネットワークを流通する多様なテキストデータの取り扱いに適するため、情報交換の基礎技術として利用されている。

近年、電子的コンテンツの著作権保護には高い関心が寄せられ、コンテンツの特性にあった様々な電子すかし技術が数多く提案されている。Web で広く流通する HTML 文書や XML 文書についても、著作権保護のための技術的手法の開発が求められている。また、ネットワーク上を流れる情報を監視して特定種類の情報の送受を発見するフィルタリング技術が発達する一方で、監視者に通信の存在自体を気付かれないように通信を行うステガノグラフィ手法も注目されている。監視の技術的有効性を考えるには監視にかかり難い通信路の構成手法を示すことも必要である。

情報ハイディング分野におけるこれまでの研究対象は主に画像・音声データへの情報の埋め込みであり、テキストへの情報ハイディング手法を取り扱った研究は多くない。特にテキストデータの構造に注目した研究の事例は少ない。

本稿では、XML 文書が構造化文書である点に注目し、XML 文書への複数の情報埋めこみ手法を 2 つのグループに分けて提案する。第 1 のグループの手法は文書構造を表現するマークアップ部分を同義の表現に変更することで秘密を文書に埋め込む。第 2 のグループの手法は、アプリケーション処理に影響を与えないように DTD で許される範囲内で XML 文書の論理構造を変更し情報を秘匿する。

以下、2 章では背景として情報ハイディングを簡潔に説明する。3 章では既存のテキスト情報ハイディング手法を紹介する。4 章以後が本論である。4 章では XML 文書に適した情報ハイディング手法を検討する。5 章では構造化文書の特徴に注目した情報ハイディング手法をより具体的に解説する。6 章で全体をまとめる。

2 情報ハイディング

この章では、背景としてステガノグラフィ手法の中心となる部分を説明し、用語を定義する。広く「情報ハイディング」として知られる研究領域としては 2 つの領域が知られている。ステガノグラフィは、一見して無害なメッセージの中に秘密を隠し、監視者から通信を不可視にする技術

*1 (株)三菱総合研究所 〒100-8141 東京都千代田区大手町 2-3-6
Mitsubishi Research Institute, Inc., 3-6, Otemachi 2-chome, Chiyoda-ku, Tokyo 100-8141 JAPAN
*2 独立行政法人通信総合研究所 〒184-8795 東京都小金井市貫井北町 4-2-1 Communications Research Laboratory, 4-2-1, Nukuikita-machi, Koganei, Tokyo 184-8795 JAPAN
*3 横浜国立大学 大学院 環境情報研究院 〒240-8501 横浜市保土ヶ谷区常盤台 79-7 Yokohama National University, 79-7, Tokiwadai, Hodogaya-ku, Yokohama 240-8501 JAPAN
*4 東京大学 情報基盤センター 〒113-0033 東京都文京区本郷 7-3-1 東京大学付属図書館内 the University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033 JAPAN

領域である。一方、電子透かしは、電子メディアの著作権保護に関する技術領域であり、管理情報を著作物自体に電子的な透かしとして埋め込む手法を扱う。これら２つは適用対象、要求される性質等の点で異なるが、メッセージを目立たないように情報中に隠す手法を扱う点では似通う。

図 1 に情報ハイディング(ステガノグラフィ)アプリケーションの中心部分のモデルを示す[Pfitzman1996]。モデルは、埋込、伝送、抽出の 3 処理から成る。embedded data は cover-text に隠される情報である。stego-text は埋め込み処理の出力であり、embedded data が中に隠される。cover-text は stego-text の原型となる入力である。stegokey は埋込と抽出の処理に必要な追加の秘密データである。

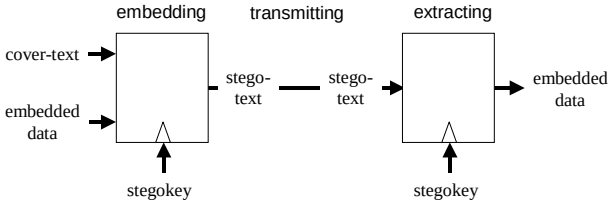

図 1：情報ハイディングアプリケーションの中心部分

3　テキスト情報ハイディング

本稿では cover-text と stego-text としてテキストデータを用いる情報ハイディング手法を指して、テキスト情報ハイディング手法と呼ぶ。以下にテキスト情報ハイディング手法の既存の研究について説明する。

3.1　テキストの内容に注目する手法

テキストを画像や文字コードとして捉えるのではなく、意味を持つシークエンスとして捉え、stego-text と cover-text の内容が同一になるように stego-text を作成する情報ハイディング手法がこのグループに該当する。

既存手法としては、同義語の置換による電子透かし手法[Nakagawa2001]、テンプレートに基づくテキスト自動生成によるステガノグラフィ手法[Chapman1997][SPAMMIMIC]などが提案されている。

内容の同一性に着目する手法は XML 文書にも応用できる。これについては 4.2 に簡潔に述べる。

3.2　印刷･表示時の整合性に注目する手法

テキストを画像的に捉えて出力時の表示/印刷形態に注目し、出力に大きな変化を及ぼさない範囲で体裁に変更を加える手法がこのグループに該当する。本質的にはテキストをイメージや印刷物として扱うものであり、多くの手法は電子すかし用途として提案されている。具体的には文字間、単語間、行間の空白の大きさやフォントサイズを微妙に変化させて情報を埋め込む手法がある。

文書の物理構造（体裁）に依存する手法は、印刷物の複写物に埋め込まれた情報が残る特徴を持つ。一方電子テキストとしては、他のフォーマット(特にプレインテキスト)への変換で、埋め込まれた秘密が失われやすい。

これらの手法の XML 文書への応用は 4.3 に述べる。

3.3　表記に着目する手法

ある電子テキストについて全く同じ内容・体裁・文書構造を表記する方法が複数ある場合は、秘密情報に応じてそれらを文書内で選択的に用いて情報を隠すことができる。既存の研究事例としては、アプリケーションで無視される行末におかれた空白文字を利用する手法[SNOW]などがある。これらの手法の多くは電子テキストへのステガノグラフィ手法として提案されている。埋め込まれた情報は表記を整えるようなフィルタ処理により失われる可能性が高い。

XML 文書の表記の特徴を活かした手法が構成できる。概要を 4.4 に述べ、より具体的には 5 章で提案する。

3.4　中間的手法

以上の 3 つの手法群の中間に属する手法も存在する。例えば、テキスト中の改行位置の変更による埋め込み手法[Takizawa2001]は印刷時の整合性と表記の両方に注目する手法と言える。

4　XML 情報ハイディングの検討

構造化文書である XML に適した情報ハイディング手法を検討する。特に文書の論理構造に着目した埋め込み手法は、これまでの研究事例には見られないが、XML 等の構造化文書に適用可能な情報ハイディング手法として有望である。

4.1　XML の特徴

XML や SGML のような構造化文書は、基本的には文書には論理構造のみを持たせ、物理構造(体裁)は必要に応じて外部から付与する。XML では文書の内容(content)、構造(structure)、体裁(style)は個別に扱われ、実際の文書は複数のテキストデータを組み合わせて表現される。

内容をタグによってマークアップされたテキストを XML 文書(XML document)と呼び、文書構造を表現するための要素と属性を DTD において定義する。XML 文書のマークアップ部分の表記(representation)は文書の内容とは別に扱う。スタイルは CSS や XSL のようなスタイルシートに定義し、必要に応じて文書と組み合わせて用いる。

4.2　内容の表し方のバリエーションの利用

XML 文書中の要素の内容を同義語で表せるならば、3.1 に挙げたような内容に関するテキスト情報ハイディング手法が適用できる。同義語で表された文書がアプリケーショ

ンで全く同じ処理を受けることが前提となる。

4.3 スタイル指定のバリエーションの利用

スタイルシートを stego-text とすれば、文書の論理構造を変更せずに、物理構造に関する記述のみの変更で情報ハイディングが行える。印刷・表示された文書の見た目に関してはアプリケーションへの依存度が高いため、3.2 のテクニックを応用した手法を構成するには、想定アプリケーション環境を限定する必要がある。

4.4 表記の変更

XML のマークアップの表記の揺らぎを利用すれば、3.3 で挙げた手法を拡張して XML に特化した情報ハイディング手法が構成できる。要素や属性の冗長な表現、タグ内の空白文字の有無などは多くのアプリケーション処理で無視されるため、情報の埋め込みに利用できる。

XML 文書中で手法を適用可能な箇所は多いが、(例えば canonicalization のような)文書整形処理で埋め込んだ情報が容易に除去される性質も持つ。

マークアップの表記を利用する XML 情報ハイディング手法については 5.1 で提案する。

4.5 論理構造のバリエーションの利用

同じアプリケーション処理結果(例えば同一の表示)が得られるような、異なる論理構造を持つ XML 文書を作成可能ならば、埋め込む秘密に論理構造を対応させて情報ハイディングが行える。この手法は cover-text の論理構造を変更した stego-text を生成する。この構造化文書の論理構造に注目する手法はアプリケーションに強く依存する。

これらの文書の論理構造に注目した XML 情報ハイディング手法については 5.2 で提案する。

5 提案手法

構造化文書の特徴を踏まえ、XML 文書に対する情報ハイディング手法を以下に提案する。

図 2 に提案手法の概要図を示す。提案手法では、ある XML アプリケーションにおける XML 文書を cover-text として情報を埋め込んで、同じアプリケーションで利用可能な stego-text の XML 文書を作成することを想定した。cover-text と stego-text の両方についてアプリケーションから同じ処理結果が導けるならば妥当な情報の埋め込み手法と見なせる。

本稿では XML 文書に情報を埋め込むための基本アイデアの紹介に主眼を置き、文書の表記と論理構造に関する手法を取り上げ、内容や体裁に関する手法は扱わない。

以下では、XML 文書の表記に注目した手法のグループと、論理構造に注目した手法のグループを順に説明する。

5.1 XML 文書の文法に着目する手法

XML1.0 勧告[XML]および XML Namespaces 勧告[XMLnames]により、XML 文書に関して同じ情報を示すための複数の構文が定義されている。また、文書が示す情報の同一性は Canonical XML 勧告[XMLc14n]に示された canonical form の比較により判定できる。XML 標準の観点では同一の canonical form を持つ文書は同じ論理構造を持つため、同じ情報を示す XML 文書として取り扱う。想定する XML アプリケーションが XML 文書表記上のバリエーションに対応可能ならば、cover-text を秘密に基づいて変形して同じ論理構造で異なる表記の stego-text を作ることで情報ハイディングが行える。手法の適用に関してはアプリケーションの特定の性質に依存する部分は少ない。

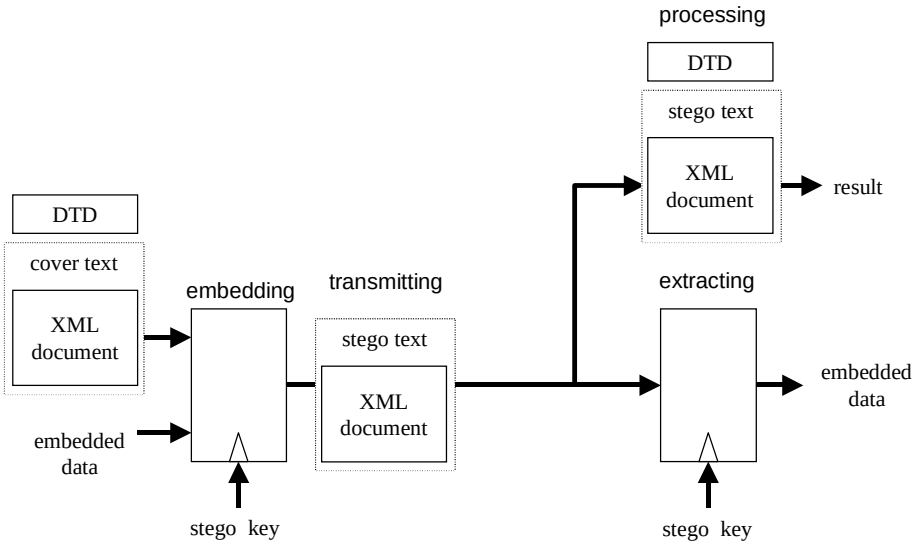

図 2：XML ステガノグラフィ

これらの手法から得られる stego-text は cover-text と同じ論理構造(同じ canonical form)を持つ。XML 文書の表記の違いに依る単純な手法であるため、XML 文書汎用の整形処理が stego-text に施されれば embedded data が消失する恐れがある。以下に手法を説明する。

5.1.1 タグ中の空白文字の利用

XML1.0 勧告[XML]によれば、表記上はタグを閉じるブランケットの前にいくつかの空白を配置でき、これらの空白は処理では無視される。これを利用して cover-text の中の各タグに、embedded data と stegokey に応じて空白を追加/削除することで stego-text を作成できる。

以下の例では閉ブランケット直前のスペースの有無により情報を埋め込んでいる。例に示す手法ではタグ 1 つごとに 1 ビットの情報を stego-text に埋め込んでいる。

Example 1.

stego key (表記:埋め込むビット)

```
<tag>, </tag>, or <tag/>     : 0
<tag >, </tag >, or <tag /> : 1
```

stego-text

```
<user ><name>Alice</name >
<id >01</id></user>
<user><name>Bob</name>
<id>02</id ></user >
```

embedded data:

```
101100 010011
```

5.1.2 空要素の表記のバリエーションの利用

XML1.0 勧告[XML]によれば、XML 文書中の空要素は、開始タグと直後の終了タグの 2 つのタグ、あるいは 1 つの空要素タグで表せる。文書中で空要素の 2 つの表記を embedded data と stegokey に合わせて切り換えることで、stego-text を作成できる。

以下の例では cover-text 内の img 要素の表記に注目している。この手法では</img>タグあるいは<img/>タグ 1 つごとに 1 ビットの情報を stego-text に埋め込んでいる。

Example 2.

stego key (表記:埋め込むビット)

```
<img></img> : 0
<img/>      : 1
```

stego-text

```
<img src="foo1.jpg"></img>
<img src="foo2.jpg"/>
<img src="foo3.jpg"/>
<img src="foo4.jpg"/>
<img src="foo5.jpg"></img>
```

embedded data

```
01110
```

5.1.3 タグ中の属性の出現順序の利用

XML 文書中の要素内に属性が記される順序を入れ換えても文書の論理構造は変化しない。cover-text 中の注目する要素について、embedded data と stegokey に応じて属性の出現順序を入れ換えて stego-text を作成することで情報を埋め込むことができる。手法は文書中で要素に複数の属性が示されている箇所に限って適用できる。注目する要素と属性の範囲は様々に設定可能である。

以下の例では、要素 event の属性"month"と属性"date"が記される順番に注目してこれらの属性の値が指定されている場所全てに情報を埋め込んでいる。2 つの属性の出現順序を入れ換える度に 1 ビットの情報を stego-text に埋め込んでいる。

Example 3.

stego key (表記:埋め込むビット)

```
<event month="MONTH" date="DATE">
  EVENT</event>      : 0
<event date="DATE" month="MONTH">
  EVENT</event>      : 1
```

stego-text

```
<event month="JUL" date="4"> Independence
  day</event>
<event date="25" month="DEC">
  Christmas</event>
```

embedded data

```
01
```

5.1.4 属性のデフォルト値の利用

要素内の属性については DTD における属性リスト宣言でデフォルト値が指定される場合がある。このような属性の値が文書中で明示的に指定されない場合はデフォルト値が用いられる。cover-text 中で属性のデフォルト値を用いる箇所について、デフォルト値を明示するよう表現を書き換えても文書の意味は変化しない。これを利用し、cover-text において属性がデフォルト値を取る際に、値を明示する/しないを embedded data と stegokey に合わせ使い分ければ stego-text に情報を埋め込むことができる。デフォルト値以外の値が使われている箇所は埋め込み位置とみなさず、何も変更しない。

以下に例を示す。この例では DTD には要素"book"の属性"language"のデフォルト値が"english"と定義されているものとしている。

Example 4.

stego key (表記:埋め込むビット)

```
<book language="english">    : 0
<book>                       : 1
```

stego-text

```
<book language="english">foo1</book>
<book>foo2</book>
<book language="japanese">foo3</book>
<book language="english">foo4</book>
```

embedded data

```
010
```

5.2 XML 文書の論理構造に着目した手法

実際の XML アプリケーションでは、同じ意味を持つ文書として異なる論理構造を持つ文書を扱う場合もある。このような場合には、異なる論理構造の文書が記述可能であるように柔軟な(冗長な)DTD が定義されることが多い。

cover-text とは異なる論理構造を持つが同一の処理結果を返す XML 文書のバリエーションを、埋め込む情報と鍵情報に基づいて作成し、これらを stego-text とすることで情報ハイディングが行える。アプリケーションは同義と見なすが論理構造は異なる文書の例を用いたいくつかの手法を以下に説明する。

5.2.1 同名要素の出現順序の利用

XML 文書中には同名の要素が連続する箇所が多く見られる。これらの同名の要素に関して内容や属性で順位付けを行うことができれば、文書中で同名要素が一定回数連続する部分を順位の数字の並び(パターン)と見なせる。パターンに情報を対応付ける stegokey を作成すれば、cover-text の同名要素(複数)を embedded data と stegokey に合わせて並べ替えて stego-text を作成できる。

以下の例では文書中の last_name 要素が 2 連続する部分に注目している。連続する last_name 要素の内容の文字列に関してアルファベット順で数をつけ、この数の並びにより stegokey を作成している。例では、要素 2 つごとに 1 ビットの情報が埋め込めている。

この手法は standard MIDI file を埋め込み対象とするステガノグラフィ手法 [Inoue2000]で示されたアイデアに基づく。

Example 5.

stego key (表記:埋め込むビット)

```
順位1-順位2 : 0
順位2-順位1 : 1
```

stego-text

```
<last_name>Smith</last_name>
<last_name>Brown</last_name>
```

embedded data

```
1 (“S”mith と“B”rown では“S”mith の方が後)
```

5.2.2 異名要素の出現順序の利用

XML 文書中に異なる要素が同じ階層に並べられている場合に、要素が現れる順序とは無関係にアプリケーションが処理結果を作るならば、文書に出現する順序を変更しても問題とならない。これを利用して、cover-text 中の同階層に並べられた要素の順序を embedded data と stegokey に基づいて変更し、stego-text を作成できる。

以下の例では user 要素の子要素である name 要素と id 要素が出現する順序を変更し、stego-text ごとに異なる論理構造を持たせて情報を埋め込んでいる。例えば name、id の順に記述されている場合は埋め込まれた情報は 0 である。ここでは 2 つの要素の出現順序を入れ換える度に 1 ビットの情報を埋め込める。

Example 6.

stego key (表記:埋め込むビット)

```
<user><name>…</name><id>…</id></user> : 0
<user><id>…</id><name>…</name></user>: 1
```

stego-text

```
<user><name>Alice</name>
<id>01</id></user>
<user><id>02</id>
<name>Bob</name></user>
```

embedded data:

```
01
```

5.2.3 同義要素の利用

処理上は全く同義に扱われる要素が異なる要素名で複数定義されることがある。このような冗長な文書定義を XML 文書が持つ場合には、意図的な要素名の使い分けにより情報を埋め込んだ stego-text を作成できる。

以下の例の family_name 要素と last_name 要素は同じ処理を受ける要素である。文書中にこれらの要素が 1 回出現するごとに 1 ビットの情報を埋め込んでいる。

Example 7.

stego key (表記:埋め込むビット)

```
<family_name>…</family_name>: 0
<last_name>…</last_name>     : 1
```

stego-text

```
<last_name>Smith</last_name>
<family_name>Brown</family_name>
<family_name>Rose</family_name>
<last_name>Woods</last_name>
```

embedded data

```
1001
```

5.2.4 要素間の包含関係の利用

定義上 2 つの異なる要素が互いに他を内包可能なら、要素の包含関係を部分的に改変して異なる論理構造を持つ文書を作り、情報を埋め込める可能性がある。

以下の例は内側のタグと外側のタグの交換によって情報を埋め込む場合を示している。入れ換え 1 つ毎に 1 ビットの情報を埋め込んでいる。

Example 8.

stego key (表記:埋め込むビット)

```
<favorite><fruit>…</fruit></favorite> : 0
<fruit><favorite>…</favorite></fruit> : 1
```

stego-text

```
<fruit><favorite>orange</favorite></fruit>
<favorite><fruit>apple</fruit></favorite>
```

embedded data

```
10
```

5.2.5 無意味な空要素の利用

ある空要素が文書中に含まれても処理結果に影響しないならば、このような無意味な要素を文書中に付加あるいは削除して、秘密を埋め込むことが可能である。

例では何ら処理結果に影響を及ぼさない空要素とその属性を利用して情報を埋め込んでいる。この無意味な空要素 1 つにつき文書に 1 ビットの情報を埋め込み可能である。

Example 9.

stego key (表記:埋め込むビット)

```
<title type="paperback"></title>     : 0
<title type="magazine"></title>      : 1
```

stego-text

```
<title type=”paperback”>BOOK 1</title>
<title type="magazine">BOOK 2</title>
<title type="paperback"></title>
<title type="paperback">BOOK 3</title>
<title type="magazine"></title>
```

embedded data

```
01
```

6 まとめ

本稿では XML 文書に適した情報ハイディング手法を検討した。構造化テキストに特有の情報ハイディング手法として、XML 文書のマークアップ部分の表記に関連する情報埋め込み手法と、XML 文書の論理構造に関連する埋め込み手法をそれぞれ複数提案した。これらの手法は電子透かし用途よりも XML 文書を利用したステガノグラフィ手法としての応用により適するものと考えられる。

今後の課題としては、提案したステガノグラフィ手法の具体的な XML アプリケーションへの適用、埋め込み可能な情報量の検討があげられる。

謝辞

通信総合研究所非常時通信グループの大野氏、山村氏、三輪氏、横浜国立大学松本勉研究室の井上氏、赤井氏、吉岡氏ほか、貴重な御討論を頂いた皆様に感謝致します。

参考文献

[Chapman1997] M. Chapman, G. Davida, "Hiding the Hidden: A Software System for Concealing Ciphertext as Innocuous Text," Financial Cryptography, First International Conference, FC'97, pp.335-345, Feb. 1997.

[Inoue2000] D. Inoue, T. Matsumoto, "Standard MIDI Files Steganography," IEEE-PCM 2000 Conference Proceedings, pp.328-331, 2000.

[Nakagawa2001] 中川裕志,三瓶光司,松本勉,柏木健志,川口修司,牧野京子,村瀬一郎: "意味保存型の情報ハイディング--日本語文書への応用--",情報処理学会論文誌, Vol.42, No.9, pp.2339-2350, 2001 年 9 月.

[Pfitzman1996] B. Pfitzman, "Information Hiding Terminology," Information Hiding First International Workshop, LNCS(1174), Springer, pp.347-350, 1996.

[Takizawa2001] 滝澤修, 山村明弘, 中川裕志, 松本勉, 村瀬一郎, 牧野京子, 井上信吾, 大野浩之: “改行位置の制御によるテキストステガノグラフィの提案”, 言語処理学会第 7 回年次大会, C2-4 , pp.135-138, 2001 年 3 月.

[SNOW] "The SNOW Home Page," http://www.darkside.com.au/snow/

[SPAMMIMIC] "spammimic - hide a message in spam," http://www.spammimic.com/index.shtml.

[XML] "Extensible Markup Language (XML) 1.0 (Second Edition)," http://www.w3.org/TR/REC-xml, Feb. 2001.

[XMLc14n] "Canonical XML Version 1.0," http://www.w3.org/TR/xml-c14n, Mar. 2001.

[XMLnames] "Namespaces in XML," http://www.w3.org/TR/REC-xml-names, Jan. 1999.