

## 1. はじめに

コンピュータネットワークの普及に伴い、ネットワーク上で公開されたコンテンツに対する著作権の確保が大きな問題となっている。それを解決する一つの方法として注目、研究されているのが情報ハイディングである。しかし、従来の情報ハイディングは、画像、音声などに対するものがほとんどであり、テキストに対する場合でも、空白の位置をずらすなどして、画像的な扱いをするものがほとんどであった。

そこで、本研究では、自然言語処理を情報ハイディングに応用し、文書中の表現を意味が変わらない他の表現に置き換えることによって情報を隠蔽することを試みた。本稿ではその概要を報告する。

## 2. テキストへの情報ハイディングシステム

### 2. 1. 文書中の表現の置き換え

日本語に限らず、自然言語には同一の内容であっても、複数の言い回しが存在することが多い。また、日本語にはさらに、漢字、平仮名の混在による表記の揺れが存在する。従って、同一の内容を表す語への置き換え、漢字の平仮名への変換、又はその逆を行っても、意味の変化がなく、且つ文書が不自然なものとならないと考えられる。

### 2. 2. システム概要

本情報ハイディングシステムの方式や特徴を以下

に述べる。

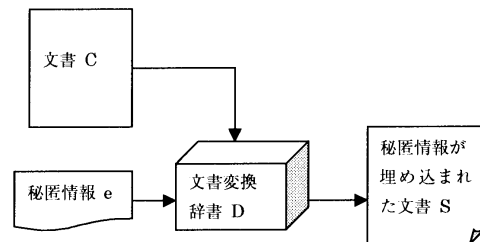


図 1. 情報隠蔽の流れ

図 1 に示すように、我々の開発したシステムでは辞書を用いて情報隠蔽を行う。あらかじめ置換対象となる語と、置換候補の語およびその候補に置換可能であるときの条件、そしてその候補に対するビット情報を保持する辞書 D を用意しておく。この辞書を利用し、文書 C 中の置換可能である部分を特定し、バイナリストリングとして入力された秘匿情報 e に対応する置換候補で置き換える。

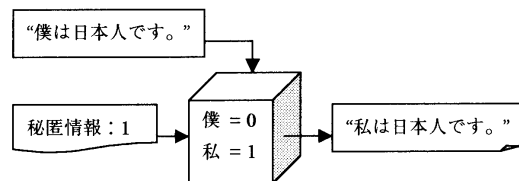


図 2. 情報隠蔽の例

実際のシステムでは、文書 C と辞書 D を用い、置換可能な部分にタグ付けを行った中間形式のファイルを作成してから情報隠蔽を行う。情報抽出には、このファイルを用いて情報隠蔽箇所を特定するという方法をとっている。

### 3. 検証実験

#### 3. 1. 意味同一性の検証

同程度のサイズを持つ、ソフトウェアのマニュアル(文書 A、B)と使用許諾文書(文書 C、D)、それぞれを 2 種類ずつ用意し、登録内容の異なる辞書 2 つ(辞書 1、2)を用いて各々に対して情報隠蔽の操作を行った。それによって作成された 8 種類の文書を被験者に査読させ、元の文書との意味の比較実験を行った。

使用した辞書、被験者によって評価にばらつきが出たが、全体的評価の平均は高く、意味がほぼ保存されていることが言えた。

①	②	③	④	⑤	⑥	⑦	⑧
A1	5	50	48	23	1	2.1%	3
	5	50	48	23	0	0.0%	5
A2	5	17	17	10	0	0.0%	4
	5	17	17	10	0	0.0%	5
B1	5	310	304	123	32	10.5%	5
	5	310	304	123	32	10.5%	4
B2	6	37	32	16	3	9.4%	5
	6	37	32	16	4	12.5%	2
C1	6	76	74	30	1	1.4%	5
	6	76	74	30	0	0.0%	5
C2	5	52	48	26	0	0.0%	5
	5	52	48	26	0	0.0%	5
D1	4	154	153	86	4	2.6%	5
	4	154	153	86	9	5.9%	4
D2	4	39	32	15	0	0.0%	5
	4	39	32	15	8	25.0%	1

- ① 文書名
- ② 文書サイズ (KBytes)
- ③ 埋込可能個所数
- ④ 埋込bit数
- ⑤ 表現置換個所数
- ⑥ 意味不一致指摘個所数
- ⑦ 不一致指摘率
- ⑧ 全体評価 (1:低い 5:高い)

図 1：意味同一性検証実験

#### 3. 2. 文書の正当性の検証

前述と同様の情報隠蔽操作を行った文書を用意し、被験者にその文書を査読させ、その文書が日本語として適切、かつ、専門用語等の用法が適切であるかを評価してもらった。

意味同一性実験と同様、辞書、被験者で評価にばらつきが出た。また、表現を置換した部分以外の個所が、不適切であると指摘されている場合が見られた。これは表現の置換が、周辺の表現にも影響を与えることによるものと思われる。

①	②	③	④	⑤	⑥	⑦	⑧
A1	5	50	48	26	5/12	19.2%	2
	5	50	48	26	3/3	11.5%	5
A2	5	17	17	10	4/14	40.0%	3
	5	17	17	10	0/0	0.0%	5
B1	5	310	304	151	83/98	55.0%	2
	5	310	304	151	13/20	8.6%	3
B2	6	37	32	17	6/12	35.3%	4
	6	37	32	17	6/32	35.3%	3
C1	6	76	74	39	8/10	20.5%	4
	6	76	74	39	11/14	28.2%	4
C2	5	52	48	26	1/3	3.8%	3
	5	52	48	26	2/5	7.7%	1
D1	4	154	153	67	9/12	13.4%	5
	4	154	153	67	28/34	41.8%	4
D2	4	39	32	14	2/2	14.3%	5
	4	39	32	14	4/9	28.6%	4

- ① 文書名
- ② 文書サイズ (KBytes)
- ③ 埋込可能個所数
- ④ 埋込bit数
- ⑤ 表現置換個所数
- ⑥ 問題指摘個所 (置換個所/全体)
- ⑦ 問題指摘率
- ⑧ 全体評価 (1:低い 5:高い)

図 2：文書正当性検証実験

### 4. まとめ

我々の提案した方式は、適切な辞書を用いれば、日本語として自然で、内容が適切であるテキストが生成されることがわかる。

現段階では、ある表現に対する置換候補、及び置き換えの際の条件設定はすべて人手で行われているため、置き換えにおける言語学的論理の解明、および、そのシステムへの反映が今後の研究課題となる。

### 謝辞

本研究は情報処理進事業協会のプロジェクトの一環として行われた。

### 参考文献

- [1] 株式会社三菱総合研究所 「インフォメーションハイディングの技術調査」報告書, 1998年2月.
- [2] SH Low and NF Maxemchuk and J.Brassil and J.O'Gorman. Document Marking and Identification using Both Line and Word Shifting *Infocom 95, April 1995.*
- [3] 小俣 祐介 テキストへの情報ハイディングに関する研究, 横浜国立大学修士論文 1999年