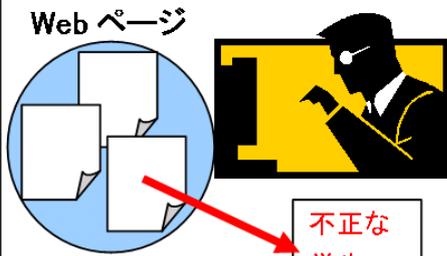


Web ページをコピーして作成された不正な学生レポートの発見支援

北里大学 一般教育部 高橋 勇



Web ページ

コピー＆ペースト
一部の書き換え

不正な
学生
レポート

不正発見戦略

- ・キーワード抽出
- ・Web検索
- ・類似度評価
- …など

一括
調査

証拠
資料
作成

不正発見の負担軽減
学生指導の支援へ

レポート	状態	評価
Data0.txt	調査完了	64%
Data1.txt	調査完了	34%
Data10.txt	調査完了	10%
Data11.txt	調査完了	48%
Data12.txt	調査完了	17%
Data13.txt	調査完了	15%
Data14.txt	調査完了	21%
Data15.txt	調査完了	14%
Data16.txt	調査完了	26%

URL	評価
TYPEAhttp://www.atmarkit.co.jp/network/tokusyuu/1310gbe/10gbe04.html	64%
TYPEAhttp://futaechan.blog106.fc2.com/blog-date-200709.html	26%

検索結果詳細

http://www.atmarkit.co.jp/network/tokusyuu/1310gbe/10gbe04.html

10ギガビット・イーサネットの標準化が「IEEE802.3ae」として承認された。これまで標準化に先行して製品を出荷してきたベンダに加え、年内にはさらに多くのベンダが標準に準拠したさまざまな製品の出荷を開始する。現状としては、先行出荷の製品を含め、レイヤ2/3スイッチにLAN PHYを実装したものが中心となっている。その用途として、ギガビット・イーサネットに代わる大容量のレイヤ2/3スイッチ間接続を意識しているのではないだろうが、ルータでは、現状でもかなりの高価格のOC-192PacketoverSONETインターフェイスの代替が期待されている。

多くの教員に使ってもらえるシステムへ

論文精読まとめ

・文献情報

深谷 亮, 山村 毅, 工藤 博章, 松本 哲也, 竹内 義則, 大西 昇, 単語の頻度統計を用いた文章の類似性の定量化—部分的類似性の考慮—, 電子情報通信学会論文誌 VOL.J87-D2 No.2, pp.661-672, 2004

・文献選択理由

議論の幅を広げるために、自分がしている教育システムの研究には関わらなくても、直接的には教育支援システムの研究ではない技術的な基礎研究にあたる文献をあえて選んでみました。

・文献が掲げる目的

オリジナルな文章と他人の文章を真似て作成した文章を区別する方法を明らかにすること（特に既存の文章の単語を置き換えるような真似に対応可能な文章類似性評価の手法を提案すること）。

・目的達成のための手段

まず、他人の文書を真似て書かれた文章の分析と類型化により研究対象の明確化をしている。そのうえで、単語間の類似性を評価する方法としてEDR概念辞書を用いた評価式を提案し、文章を文へ分割する手法と形態素解析を用いて、文章を構造をもつ単語の集合に置き換え、これを用いて文間および文章間の類似性を評価する式を定義している。この評価式により、文章の類似性を定量化している。

・評価の方法

学生にオリジナルな文章と真似をして作った文章とを作成させ、提案する評価式を用いて類似性を定量化し、他の定量化の手法（n-gram及びTF-IDFの方法）と比べてどのくらい両者を区別できるかを評価している。また、実際に提出された学生レポート間での評価も行なっている。