

自分で出来るナレッジマネジメント

- 平成17年度
福岡県・佐賀県大学図書館協議会総会
における 講演
- 2005年5月20日
- デスクトップ検索を追記しました。5月23日

自己紹介

- 西南学院大学 商学部 経営学科
教授 吉武春光(工学博士)
yositake@seinan-gu.ac.jp
<http://www.seinan-gu.ac.jp/~yositake/>
 - 情報処理センターの所長
 - 長らく、本学のインターネット環境の構築に携わった。
- 担当科目
 - 情報ネットワーク論
 - 情報処理概論
 - ビジネス情報技術入門・演習

内容

- 経営分野のナレッジマネジメント
- Adobe PDF の活用
- デスクトップ検索

経営分野のナレッジマネジメント

- 経営戦略の一環として、最近、話題になっている。
- ITを活用して情報共有を実現すること。

経営分野での目的

- BPR, ERP, SCM, CRM などの導入により、企業活動がスピードアップし、情報の共有も実現された。
- しかし、数値化されていない社員のアイデアや知識やノウハウなどが活かされていない。
- 問題
 - リストラにより知識が流出。
 - 人を育てるのが間に合わない。

- 
- そこで、知識＝ナレッジ(Knowledge)の共有
 - (ナレッジ・コラボレーション)

前提

- 1人1人が、しっかりナレッジの管理が出来ている！？
- しかし、それは本当か？

乗り越えるべきハードル

- 情報共有の主体となる人間の意識
- コンピュータ・リテラシー
- IT環境の整備

知識共有の課題

■ 表現、入力、蓄積

- 誰もが理解できる形で表現するには時間がかかる。
- 文字だけでは表現しにくい。
- 入力すべきかどうか、判断に迷う。
- 失敗事例は入力されにくい。

■ 利用

- うまく検索できない。

- 1つの形式に変換するのは面倒なので、複数の形式のまま、知識共有をしたい。

- 膨大な情報から知識を抽出する
 - 類似語検索
 - データマイニング

吉武の場合

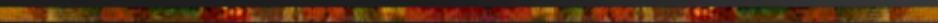
- 研究室は、本や配布文章が散乱している。
 - 先日の地震のために、更にひどく散らかっている。
- 電子メールにて、色々なデータが飛び込んでくるが、日に100通近くの電子メールを受け取るので、処理し残している電子メールが徐々に溜まってしまい、過去メールの参照が出来ていない。

超整理法

- なかなか実践できない...
- Last Used の優先順位を上げる。
- ラベル付けはしない方が良い。
 - ラベル付けが混乱することがある。
 - 電子メールをフォルダに分けて入れることで分類しているが、探し物をする時には、なかなか目的とする電子メールが見つからない。

検索

- 保管しているデータから目的とするデータを検索出来ると、便利。
- しかし、検索可能なのはテキストデータのみ。
- 画像データや紙文章は検索できない。

- 
- どうにかして、
 - ・文字テキスト
 - ・画像
 - ・紙配布物
- などを一括して管理し、検索をしたい。

画像の検索は出来ない

- ファイル名を工夫しておく。
- または
- 画面に、何か文字を打ち込んでおく。

Adobe PDF

- Adobe という会社が1993年に始めた共通文書規格
 - Windows や Macintosh などの異なるプラットフォーム間で文章をやりとりしたい。

■ Adobe Reader

- PDF ファイルを読むために必要なソフトウェア。
- 以前は Acrobat Reader と称していた。

■ Adobe Acrobat

- PDF ファイルを作成するために必要なソフトウェア

■ Adobe は PDF の規格を公開しているので、他社も PDF 作成ソフトウェアを作り販売が可能。

Adobe Acrobat 7.0

- 6.0 に比べて、起動時間が大幅に速くなっている！！
- Windows XP や Office などと同じようなアクティベーションが必要。
- Adobe Reader 7.0 も、起動時間が大幅に速くなっている！！

Adobe Acrobat 7.0 は数種類ある。

■ Elements

- 価格は5千円程度
- 作成と検索用途には十分。

■ Professional

- 価格は数万円
- OCR エンジンを搭載。
- PDF の合成などが可能。

- PDF は、一般には、「文章を配布するためのもの」と捉えられている。
- しかし...

透明テキスト付きPDF

- PDF の構造は、
画像 + テキスト(文字)
- 例えば、MS-Word で作成した文章を PDF に変換すると、MS-Word のレイアウトに沿って印刷された画像と、MS-Word の中身の文章が重なった形になる。

OCR

- オプティカル・キャラクター・リーダー
- スキャナで読み込んだ画像から輪郭に沿って文字情報を読み取るもの。
- 活字を印刷した画像から文字を読み取る精度は95%以上ある。
- 手書き文字の画像から文字を読み取る精度は悪い。

OCR につなぐスキャナ

- 解像度が高いほど、OCR の認識率は良くなる。しかし、解像度が高いほどファイルのサイズが大きくなるので、現実的ではない。
- 例えば、A4サイズをカラーの 600dpi でスキャンすると、サイズは 100MB を超える。
- OCR 用途であれば、カラーではなく「グレー」にてスキャンすると、ファイルサイズは 1/3 になる。
- 「白黒」にてスキャンすると、スキャン時に、文字がかすれたようになるので認識率が悪くなる。
- 「グレー」の 300dpi から 600dpi 程度が適度...?

電子スクラップ術

- 紙文章を OCR で読んで、それから直接 PDF ファイルを出力するソフトウェア
- 例えば
 - エー・アイ・ソフト(株)の「pdMaker for ScanPaper」
 - ピーアンドエーの「エキスパートPDF2 Professional Win」
 - ソースネクストの「いきなりPDF Professional」

電子スクラップ専用マシン

- 富士通の ScanSnap
- 個人向けの電子スクラップ専用マシン
- 操作は簡単。
- 1枚の紙の両面を一度にスキャンしてくれ、透明テキスト付き PDF に変換してくれる。
- 但し、紙を ScanSnap に通すタイプなので、1枚ものの配布物には適しているが、製本した紙には使えない。

Acrobat の検索機能

- 指定したフォルダの中にある PDF ファイルを検索対象にする。
- 検索時に指定できるのは、単語1個のみ。
- 絞り込みは、画面下の専用メニューから行う。
- 検索の度に、PDF の内部を検索するので時間がかかる。

- パソコンにて頻繁に書き換える Word 文章や PowerPoint ファイルなどを一々、透明テキスト付きPDF に変換するのは面倒。



- パソコン内の色々なファイルの中身を検索するソフトウェアが出現した。一般にデスクトップ検索 (Desktop Search) と呼ばれています。

デスクトップ検索 (Desktop Search)

- Google Desktop Search
- Copernic Desktop Search
- MacOS X Tiger の Spotlight
- Microsoft の MSN Search Toolbar with Windows Desktop Search
- ジャストシステムの ConceptSearch
- クセロの ePware

- デスクトップ検索は、予め、インデックスを作成しておき、そのインデックスを検索するので、検索は高速。
- インデックス作成の時間が必要。
 - パソコンを操作していない時に、勝手にインデックスを作成してくれる。
- インデックスの更新のタイミングが問題になる。

Google Desktop Search

- <http://desktop.google.com/>
- 英語版は正式版ですが、日本語版はベータ版です。
- PDF ファイルの内容も検索できるはずなのだが、日本語の検索は何かおかしい。
- 英単語の検索は正常。

Google Desktop Search の問題と制御

- 次のページが、よくまとまっている。

http://users.tns.net/%7Eskingery/firefox/GDS_Tips.html

- 問題2つ

- ネットワークドライブの指定が出来ない。
- 削除されたファイルやフォルダがキャッシュに残ったままになる。

ReIndex

- このページで紹介してある
ReIndex for Google Desktop Search
<http://reindex.wz.cz/>
から ReIndex-1.6.zip をダウンロードする。
- C:¥Program Files¥ の下にフォルダを作り解凍する。
 - 吉武の場合、ReIndex-1.6 というフォルダ名にしています。
- 吉武の場合 zドライブ(ネットワークドライブ)と、cドライブを
検索させたいので、メモ帳を使って次の内容を入れて
〇〇.bat というファイル名で保存。
 - "C:¥Program Files¥ReIndex-1.6¥ReIndex-FULL-1.3.exe" z c
- 〇〇.bat をダブルクリックすると、zドライブと、cドライブの
インデックスを作り始める。

ジャストシステムのConceptSearch

- 使ったことがないので、不明。
- PDF の中身も検索できるようなことが書いてあるが...

クセロの ePware

- <http://xelo.jp/pdf/02soft/epware.html>
- .doc や .xls などにも検索できるようにする。
- PDF の中身の検索はバッチリ
- 欠点: ファイルを一旦、専用の場所にコピーする仕組みになっているので、管理が面倒。

Copernic Desktop Search

- <http://www.copernic.com/en/products/desktop-search/index.html>
- 英語版
- PDF ファイルの内容も検索できるはずなのだが、何かおかしい。

MSN Search Toolbar with Windows Desktop Search

- <http://desktop.msn.com/>
- 英語版のみ。
 - 但し、日本語の検索は問題ない。
- Word や Excel は、ファイル名とファイルの内容を検索可能。
- しかし、PDF は、ファイル名のみを検索可能。



- Microsoft は、新しい文書フォーマット Metro を提唱し始めた。PDF 形式の対抗馬かもしれない...?

結局...

- デスクトップ検索は、出回り始めたばかり。
- PDF の中身の検索については、まだ不十分。
- 現時点では、Google Desktop Search が、
もっとも、良さそう。
 - PDF 内の英単語の検索はOK。

e-文書法

- 正式名称は
「民間事業者等が行う書面の保存等における
情報通信の技術の利用に関する法律案」
及び
「同法施行に伴う関係法律の整備等に関する
法律案」
である。
- e-Japan戦略Ⅱ の一環

- 民間企業に紙での保存が義務付けられている財務や税務関連の書類・帳票を、電子データとして保存することを認める。
- 2005年4月から実施

- 1) スキャナ方式による保存
 - OCR技術の導入
 - PDF 形式とは言っていない。

- 2) タイムスタンプ技術の適用
 - いつの時点で作成され、それ以降現在に至るまで改ざんされていないことを証明する目的