

位置情報とデータマイニング

2007.05.30

GIS総合研究所 国司輝夫

データマイニングとは

データマイニングとは、統計学、パターン認識、人工知能等の技法を大量のデータに適用することで知識を取り出す技術。

過去に蓄積されたノウハウを体系化し、知識を共有活用する「ナレッジマネジメント」や「One to One」や「CRM」などが前提の背景。

「データを知識としてより戦略的に活用すること」つまり、データマイニングが社会的に求められた。

データマイニングの利用

データマイニングは大量のデータから有用なデータのみを取り出す技術であり、うまく利用すると大変有用な技術。

大福帳のような大規模なデータベースを作成し、データを集めてから解析する方法が多い。

ただし利用の仕方によっては個人の情報の突合せなどの運用も多い。(過去の経歴検索、行動履歴分析など)

データマイニングが利用される要件

1. マイニングされた知識が実行可能である。
2. 利用者が理解できる仕組みで表現されている。
3. マイニングされた知識を人間が強化できる。
4. マイニング知識が容易にメンテナンスできる。
5. 新しいデータに対応した知識を抽出できる。

Webにおけるデータマイニング

Web上のデータマイニングにおいては検索サービスが大きなキーとなっている。

GoogleやYahooなどの検索でデータを探す人が大半で独自に探す人は少ない。

メタタグとは

HTMLに<meta name=“keywords” content=“パソコン,直販”>などとしてキーワードや内容を記入。

昔の検索エンジンは、主にこのkeywordsを読んでそのサイトの内容を定義付け。

しかしメタタグを大量のキーワードで埋め尽くし、検索エンジンを惑わす手法を濫用する人が現れたので役に立たなくなった。

SEOの登場

SEOとはSearch Engine Optimizationの略でSEO業者はある特定の検索エンジンを対象として検索結果でより上位に現れるようにWebページを書き換えるアドバイスをし始めた。

現在もSEO企業は多数存在するが検索サイトが露骨な検索エンジン用加工のページをランキングからはずすため、変則的な手法はとりにくくなっている。

Link Farmsの出現

グーグルなどの検索エンジンは、どれだけ多くの人がWebページに対してリンクを張っているかを、客観的な重要度として検索結果の表示順位の指標のひとつにしている。

Link Farmsとは検索エンジンの検索結果で特定のWebページの表示順位を上げるために、そのWebページへの不自然なリンクが大量に張られている状態。

Cloakingの出現

検索エンジンに最適化されたページの作成

人間に見えるページと異なるページを検索エンジンSpiderに提供する方法

Cloakingは特定のWebサイトのコンテンツを検索エンジンを誤解させる試み

利用者のIPアドレスなどにより、別のページを見せる手法で、特定地域の人だけに特定ページを見せるなどの使い方がある。(Geotargeting)

Geotargetingとは

Webサイトを訪問した時に、アクセス元のIPアドレス情報を通じてその利用者の位置情報をおおまかに特定すること。

グーグルアース はIPアドレスを使わない位置情報特定ツールのひとつの代表例である。

位置情報によるマイニング

これからはGPSなどの位置情報による個別のマイニングが進むと予想される。

Real Time Data Mining(業界用語?)もこれからの進展が予想される。

Realtime Recommendation(後述)も進む。

Realtime Recommendationとは

ユーザの情報を統計的に分析し、自動的に顧客の嗜好を予測するシステム。

多くの顧客のアクセス履歴からユーザと似た嗜好を持つ履歴を抽出する「協調フィルタリング」

コンテンツに属性を付加して顧客の属性とコンテンツの属性を照合する「コンテンツベースフィルタリング」

Semantic Webとは

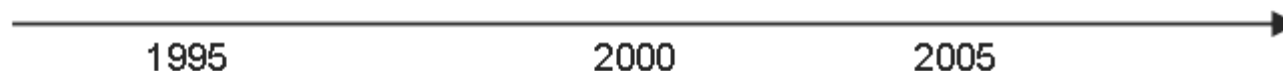
Semantic Webとは、意味を持つWebで内容に関する情報(メタデータ)をWebページでコンピュータが内容を理解できるようにし、情報を自動的に処理させる技術。

この技術が実現すれば、検索結果の精度が高まったり、Web上にあふれる情報を利用者が簡単に活用できるようになる。

2006年の前半から、日本でもSemantic Webの利点や効果に着目する人が多くなって来た。

SemanticWebへの進化

	Static	Dynamic	Syntax	Semantic
Encoding	HTML	+ RDBMS	+ XML	+ RDF/OWL
Creation	Manually	Generated by server-side applications	Generated by applications based on schema	Generated by applications based on models
Users	Humans	Humans	Humans and applications	Humans and applications
Paradigm	Browse	Create/Query/Update	Integrate	Interoperate
Applications	Browsers	Browsers	Process Integration, EAI, BPMS, Workflows	Intelligent agents, Semantic engines



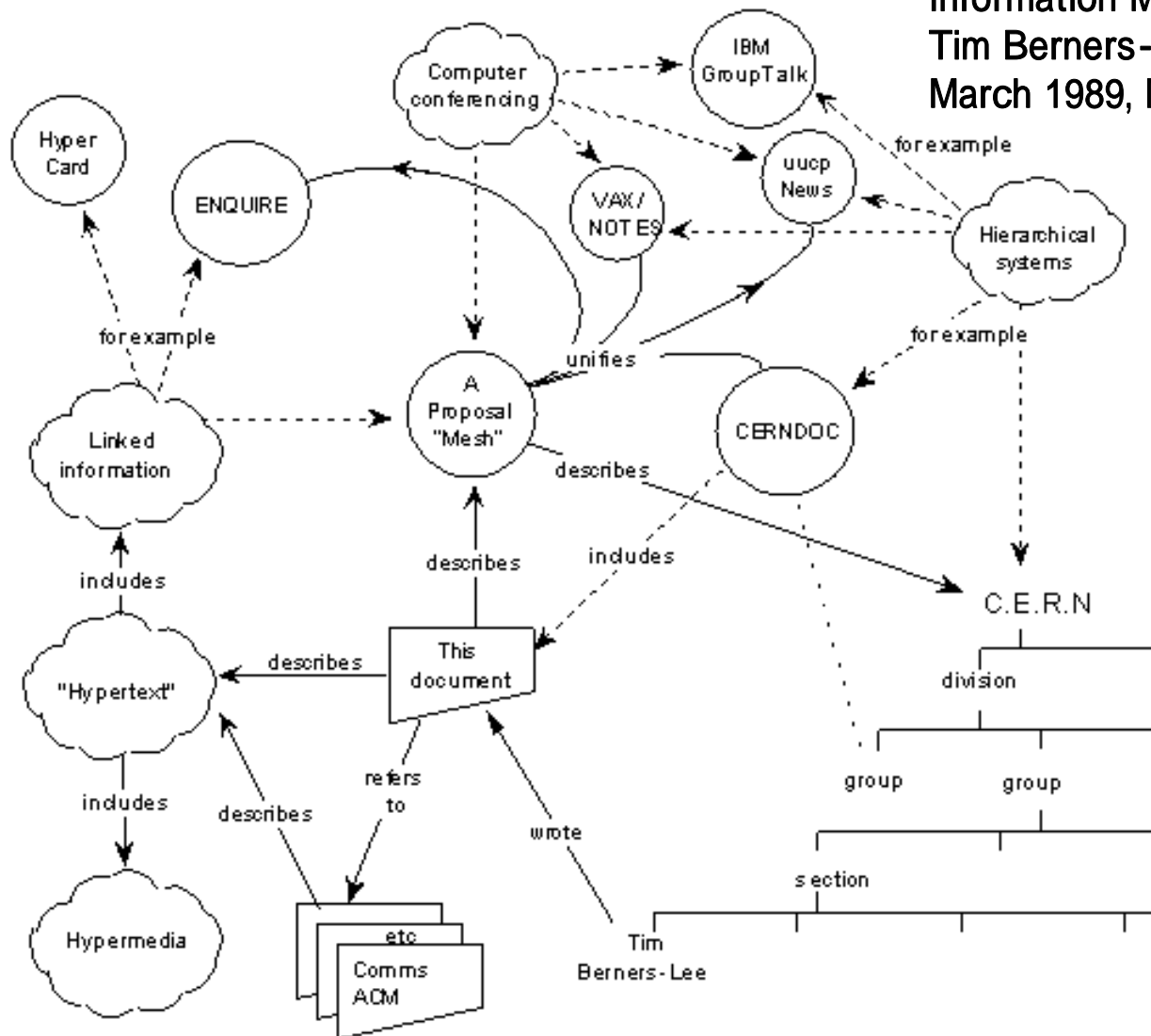
Semantic Web Services Theory Tools
and Applications Mar 2007

SemanticWebに求められるもの

1. 世界中のあらゆる情報を
2. 簡単な問い合わせで
3. 的確に探し出し
4. 要望に応じた形で
5. わかりやすく答えてくれる

SemanticWebの原点

Information Management: A Proposal
Tim Berners-Lee, CERN
March 1989, May 1990



SemanticWebとWeb2.0

Web2.0 では、ブログやソーシャルタギングによる集合知の利用

Ajax やマッシュアップによる軽量プログラミング、ロングテール現象による新しい需要

Ajax (Asynchronous JavaScript + XML)

ソーシャルタギング(フォークソノミー)とはタグと呼ばれるキーワードで直感的、多次元的に分類

SemanticWebはXML？

SemanticWebの考え自体は広く普及しているが、XML 文書として積極的に利用しようという考え方はあまり普及していない。

XML の考え方は機械的処理には非常に有用な
のあるが、人にとってはなじみにくい。
そこで、XML を使わずに HTML の範疇で
SemanticWebに近づける試みも現れている。

中間処理を加えたXMLジェネレーターのような形
に落ち着く可能性もある。

コミュニティに知らぬ間に参加？

Web2.0の世界ではコンテンツの自発的な連携と共有によってネット全体が一つのコミュニティとして機能するため、「特定のコミュニティに参加する」という発想がなくとも参加している。

ソーシャルタギングやフォークソノミー により実現

例：はてな <http://www.hatena.ne.jp/>

Flickr <http://flickr.com/>

del.icio.us <http://del.icio.us/>

新しいネットコンテンツの傾向

新しいネットコンテンツに共通しているのは、他人に見せることが第一の目的ではない。

自分のために制作、収集、整理した情報(コンテンツ)を、他人にも一部公開することで連鎖を繰り返している。

変化の本質は「技術」よりもむしろ「ユーザー側の使い方」にある。

現在地情報の活用

ユーザのコンピュータ上での操作内容から
ユーザの興味や嗜好を取得し、情報を選択する。

よく行く場所、よく使うものをキーワードに使うこと
で、効率の良いフィルタリングができる。

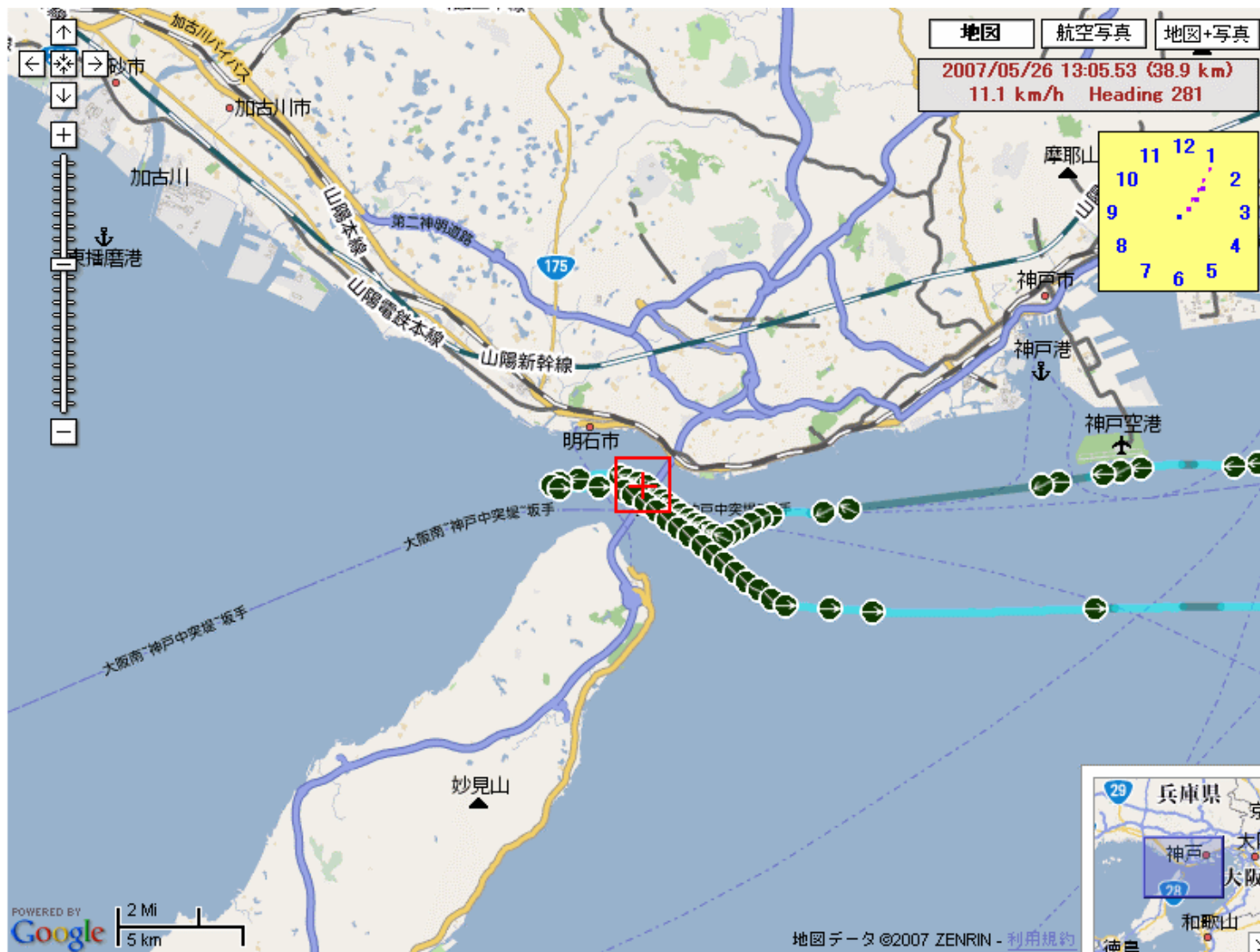
ユーザの日常行動をセンシングし、よく使う物や
よく行く場所を見つけ出す。

位置と滞在時間を使うことで、ある程度行動を特
定できる。

位置記録の例1



位置記録の例2



sDSCF0283.JPG
 2007/05/26 13:05



sDSCF0284.JPG
 2007/05/26 13:05



sDSCF0285.JPG
 2007/05/26 13:06



sDSCF0286.JPG
 2007/05/26 13:06



sDSCF0287.JPG
 2007/05/26 13:06

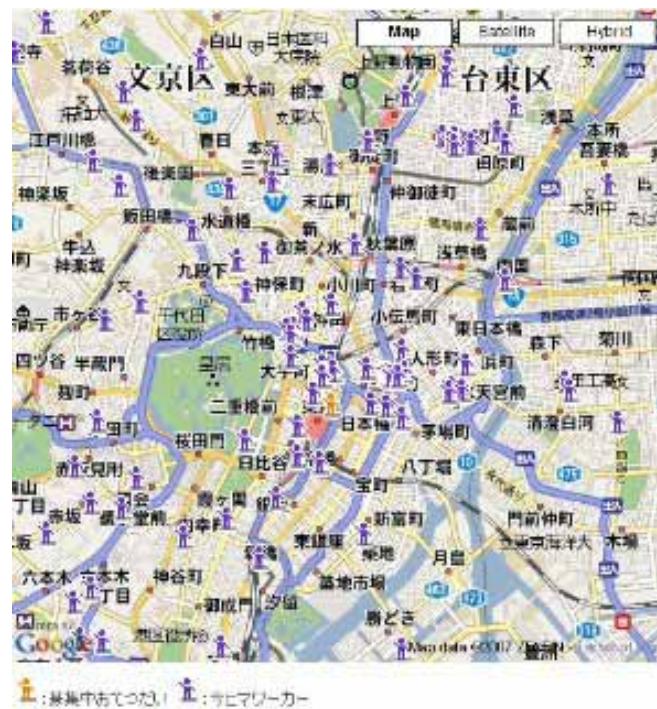
特定非営利活動法人 NPO (内閣府認証) ⑧

現在地情報の活用例1

短期雇用のマッチングをサービス
「おてつだいネットワークス」

ユーザが「今暇です」と手を挙げ
るとそれが地図上に表示される。

1時間とか2時間と言った非常に
小さな単位でも依頼することができる。



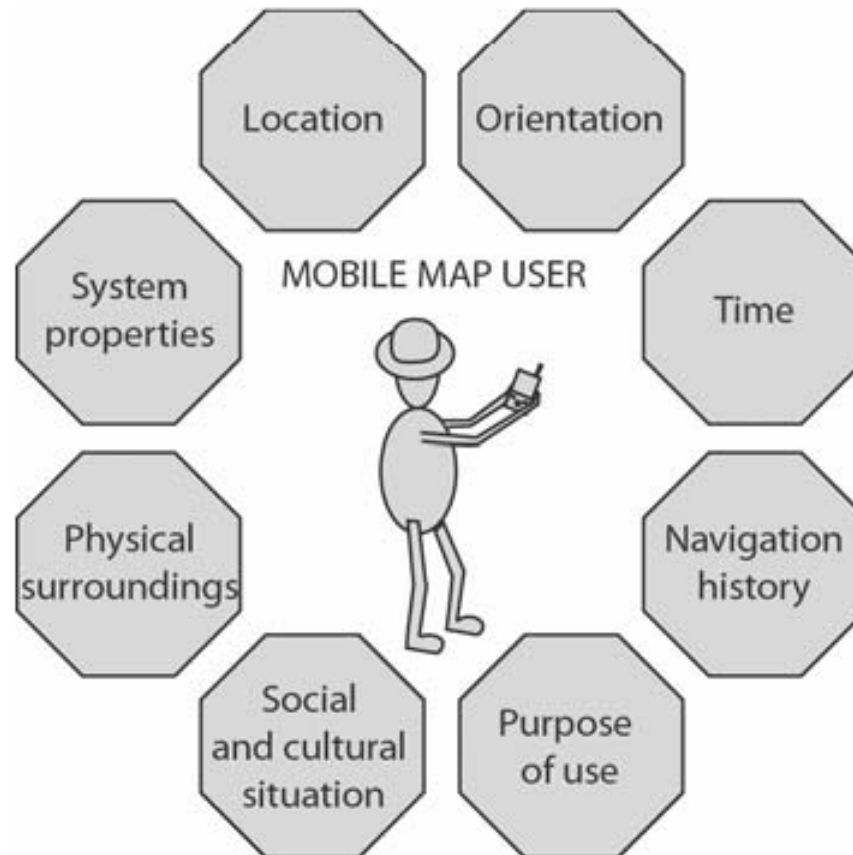
現在地情報の活用例2

災害現場での救済時に被災者の確認などが期待されている。

交通情報の渋滞や事故回避などの利用が期待されている。

利用者の嗜好と位置による店舗案内なども期待されている。

現在地情報の活用図



Map-based Mobile Services 2005

現在地情報の活用のこれから

移動履歴

「今どこにいるか」だけではなく、「何のためにここにいるか」というのを理解する。

独自商品開発

位置情報をベースにして商品を開発する。
危ないところには行かない人は保険料が安い
GPS連動保険など。

ソーシャル

「もの」や「店」とユーザとのつながりも含めて考え、その関係性の動的な変化に着目。

新しいWebマップサービスの特徴

- オープンソースベース
 - 費用がかからない環境整備
 - 利用できるリソースが多い
- インターネットと親和性
 - 簡単で共通のインターフェイス
 - インターネットの情報利用が可能
- Ajax技術、APIを提供
 - レスポンスが速い。特殊なハード・ソフトが必要ない。
- データがフリー
 - 高精細な画像、詳細なデータが無償
- ユーザ参加型
 - 技術者らのコミュニティ、ユーザコミュニティによって支持される
- 常時更新
 - ソフトウェアのバージョンアップで利益を得るのではなく、良質なサービスを提供することで、ユーザをひきつける。

参考文献

Semantic Web Technologies and E-Business
Toward the Integrated Virtual Organization and
Business Process Automation ISBN 1-59904-
194-4

Map-based Mobile Services Theories, Methods
and Implementations ISBN 3-540-23055-6

Information Management: A Proposal
Tim Berners-Lee, CERN
March 1989, May 1990
<http://www.w3.org/>