



Department of Medical Informatics
Graduate School of Medicine
Kyoto University

医療における言語処理とライフログ



京都大学大学院医学研究科 医療情報学講座
(京都大学医学部附属病院 医療情報部/運営企画室)

竹村 匡正

MTA2009 千葉大学 2009/09/13



今日のお話

- 京大病院における自然言語処理研究の紹介
- ライフログの足音
- 医療情報の現状（と今後）



背景

医療の情報化

病院情報システムの実用化

地域医療連携の促進

WWWの爆発的普及

病院内外における医療情報の爆発的普及

多メディアへの適応

データベース、画像、映像、テキスト情報



入力インタフェースとしてのフリーテキスト

再利用を考えると・・・

データベースを構築

- ・ Fill in blank入力はそれほど効率がよい訳ではない
- ・ 選択肢を選択できるということは、全ての選択肢を知る必要がある
- ・ 過剰な標準化・標準化のための標準化

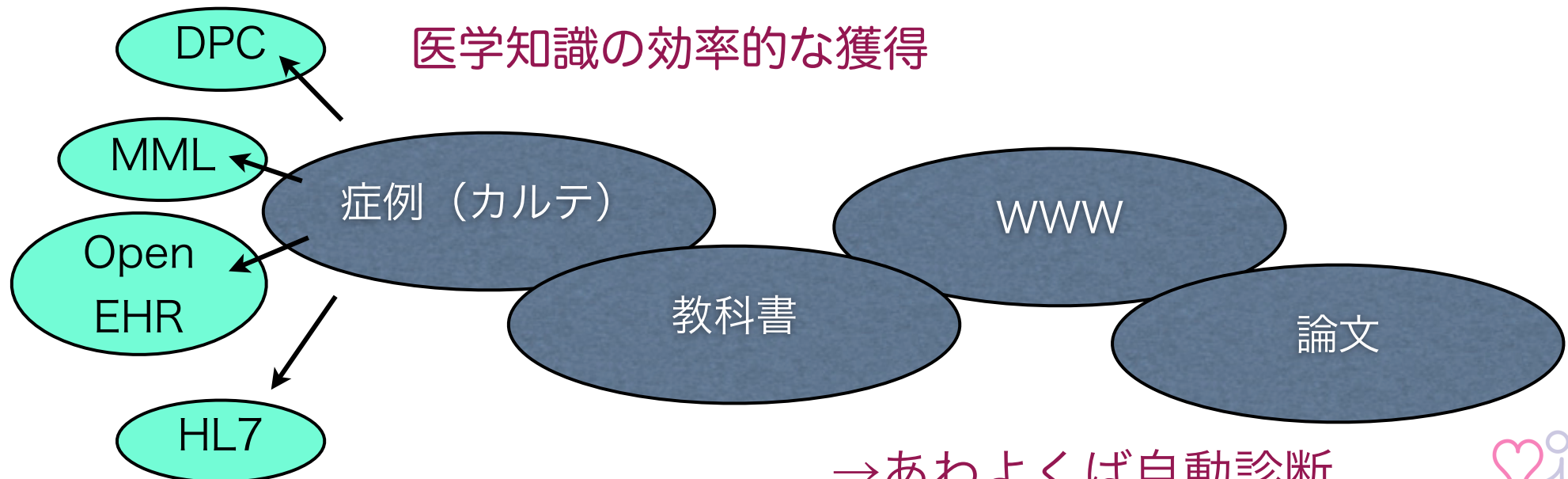
入力のし易さを考えると・・・

- ・ 非定型な情報ほどフリーテキスト（うまくDBと組み合わせ）
- ・ 知識をデータベースで表現しうる範囲は意外と少ない
- ・ WWWはテキストベース、カルテ記載もほとんどテキスト

目的

- フリーテキストで記載された知識から、多角的に様々な知識を抽出できればうれしい

類似症例検索等の意思決定サポート
医学知識の効率的な獲得



→あわよくば自動診断



言語解析

- 単語・辞書
- 形態素解析
- 係り受け解析
- 構文解析

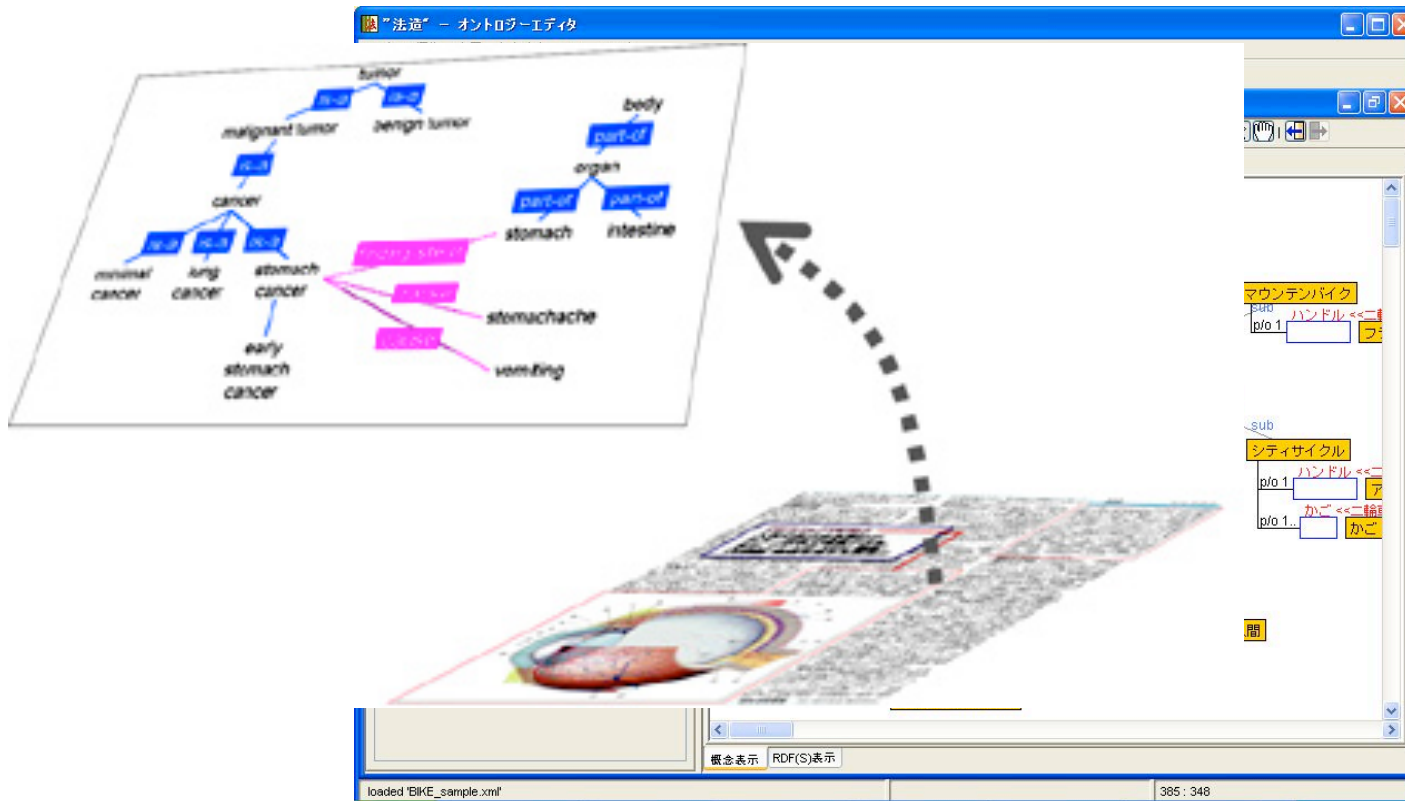
42	尺骨神経	イリヨウ解剖	尺骨神経	名詞-一般
43	、	記号-読点		
44	正中神経	イリヨウ解剖	正中神経	名詞-一般
45	、	記号-読点		
46	橈骨神経	イリヨウ解剖	橈骨神経	名詞-一般
47	、	記号-読点		
48	筋皮神経	イリヨウ解剖	筋皮神経	名詞-一般
49	、	記号-読点		
50	腋窩神経	イリヨウ解剖	腋窩神経	名詞-一般
51	を	ヲ	を	助詞-格助詞-一般
52	含み	フクミ	含む	動詞-自立 五段・マ行 連用形
53	、	記号-読点		
54	肩	カタ	肩	名詞-一般
55	の	ノ	の	助詞-連体化

ヒトの言語処理をモデル化



オントロジ

知識を体系化する



法造HPより

<http://ontsupport.enegate.jp/ontology/modules/tinyd2/>

<http://www.m.u-tokyo.ac.jp/medinfo/ont/>





ひとつの視点

- 本当に「ヒトの分析」結果が必要なのか？
 - 人が文章を理解するメカニズムが解明できれば、機械に知的な活動をさせられるのか？
 - たぶんyes→わかればわかるほど役に立つ
- 本当に「医学知識」の体系化が必要なのか？
 - あれば役に立つ・・・が、今の医学の体系が一つのオントロジで表現され得るかはわからない



コンピュータに知的な振る舞いをさせる

- ユーザがコンピュータに「こいつ賢いなあ」と思うのはどのような時か？

GoogleのPageRankのインパクト

「人気のサイトからリンクされたサイトはまた人気がある」
→中身を見ていない

人工無脳へのツッコミ

経験則に知性をみる



文章「理解」

現在の自然言語処理の得意分野→自動分類

文章内の特徴語を抽出し、その特徴語が多く含まれる分類にアサインする

TF・IDFヒューリスティクス

任意の単語があるページに多く出現(Term Frequency)し、その他のページに含まれない(Inverse Document Frequency)場合その単語はそのページ群において特徴的である、という経験則



類似症例検索

- 過去の類似した症例を検索することで、今後の診療の参考にする

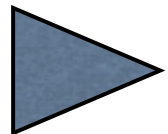
では、類似症例とは？

- 病名等ばかりではなく、診療のプロセスが近い症例（とする）



文章の「流れ」の理解

- ・ 記載内容を「医学知識」から同定する
 - ・ オントロジの利用
- ・ 書き手の「書き方」から同定する
 - ・ 客観的な文章の記載には、ヒトの記載のための知識(コンテキスト)が介在している



「診断/治療サイクルに基づく文章記載」

Observation/Diagnostic/Therapeutic (ODT) Cycle

「状況に基づく文章記載」

Context Based Structure(CBS)



文章の流れの把握

文に対して、O/D/Tのラベルを人手で付与

- T: 入院時より、
- T: すでにリマチルを含むすべての薬剤はOFFになっていた。
- O: DOEの増悪、
- OD: 胸部レントゲンの陰影の増悪がリマチル内服時に一致していたことより
リマチルによる薬剤性肺炎を疑った。
- O: リマチルのDLST,
- O: パッチテストでは陰性であった。
- ...

人がふった ODTラベルを機械学習 (SVM:Support Vector Machine)

▶ 文章の流れを自動的につかむ

流れをつかむことができると・・・

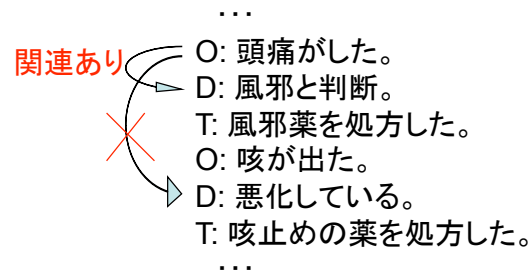


文章どうしの比較

流れを含めた文章どうしの比較→類似症例の検索

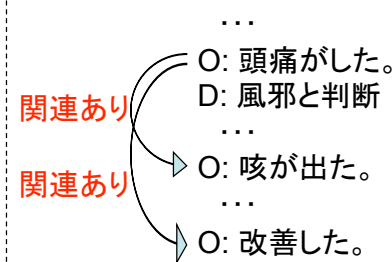
強い前後関係

1サイクル以内(ODTサイクル)



同じ分類

全体の
流れとして



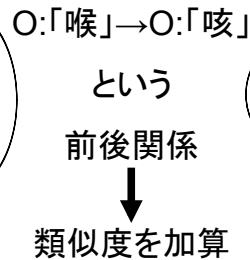
文書A

O: 喉が痛いとの訴え。
 頭痛もあり。
 D: 風邪と判断。
 T: 薬を処方し、睡眠を取る
 ように指導。
 O: 咳が続くようになった。
 ...

文書B

...

O: 喉に違和感がある。
 D: ...
 ...
 O: 咳が出たが、収まった。
 ...



類似度の高いものを類似文書とする

(岡本、竹村、2005)





類似症例の抽出

退院時サマリを対象として、ODTモデルとベクトル空間モデルで比較（そもそも抽出できるのかも含めて）

1. 同じ病名のサマリを抽出することができるか？

5病名（2型糖尿病、肺癌、狭心症、子宮筋腫、脳梗塞）、20症例ずつで比較

正解率：ODT: 72.0%, ベクトル空間: 77.6%

2. 同一疾患のサマリの中で類似症例を抽出することができるか？

肺癌サマリ22例に対して、医師の判断と比較

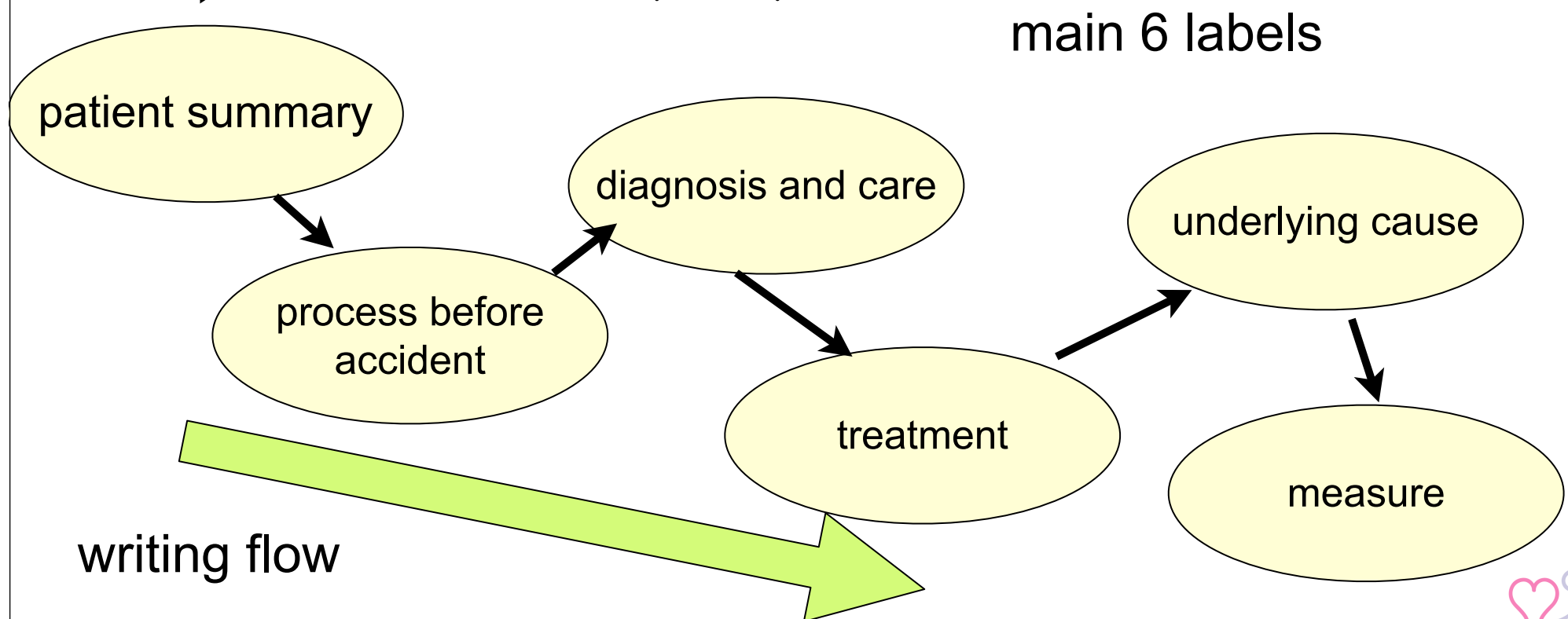
（ランキング方式：1位3点、2位2点、1位1点）

得点：ODT: 36点, ベクトル空間: 22点

良い文章、悪い文章

情報に過不足のない「良い」インシデントレポート:

▶ 典型的な書き方(CBS)が存在





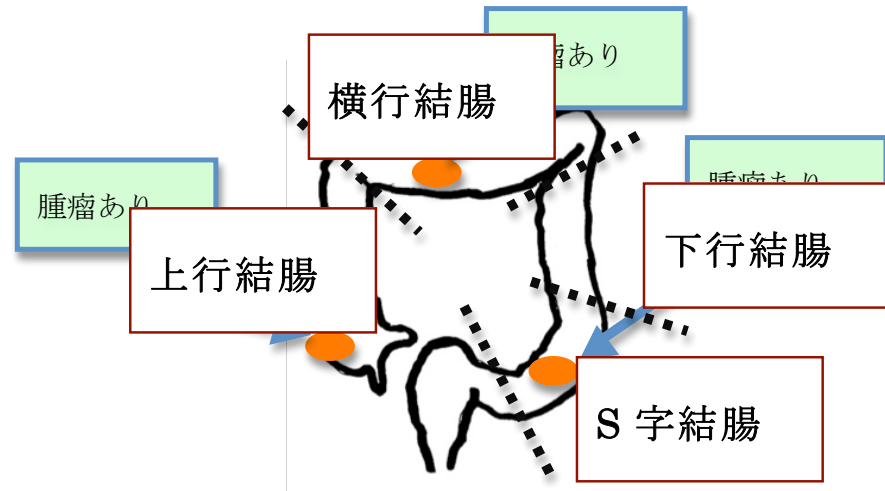
内容の自動判断/ナビゲーション

CBS label	Correct answer	correct class	precision	recall	Num. of base data
patient summary	76	125	0.61	0.35	216
process of before accident	269	362	0.74	0.60	447
diagnosis and care	244	290			
treatment	149	185			
underlying cause	10	36			
measure	122	180			

内容の判断/次に記載
 する内容をナビゲート

シエーマの位置情報の利用

- 画像+文字情報の組み合わせ
- 画像のセグメントにメタ情報を持たせる
- 類似症例検索





あいまいな医療情報の私

- 今までの医療情報（学）
 - 何とか診療に資する情報を獲得したい
 - 何とか病院運営/医療社会に資する情報を獲得したい
 - 使える形で獲得したい→構造化
 - 利用者に優しい形で獲得したい→マルチメディア/ヒューマンインタフェース
 - 情報を入力させて、HIS内のデータをリッチにする

そんなばあいじゃないかも



ライフログの足音

- 生活 (Life) 記録 (Log)
 - 人間生活の全ての行動を記録する
 - 記録したものをどうするか・・・はわからない
 - コンセプトとしては目的志向型ではない、とってみて考える (情報? データ?)
 - 個々に目的化したサービスが展開されつつある



研究レベル

上野公園1, 東京都文京区本郷 天気:晴れ 2005/07/11 12:58:47 ↓ 13:54:20	上野公園2, 東京都台東区上野公園 天気:晴れ 2005/07/11 13:33:01 ↓ 13:40:54	上野公園3, 東京都台東区上野公園 天気:晴れ 2005/07/11 13:42:58 ↓ 14:32:32	上野公園4, 東京都台東区上野公園 天気:晴れ 2005/07/11 14:09:41 ↓ 14:25:47
寛永寺, 東京都江東区豊洲見 天気:曇り 2005/07/11 17:03:41 ↓ 18:07:35	駒込, 東京都文京区本郷 天気:晴れ 2005/12/29 16:40:09 ↓ 17:12:08	両国1, 天気: 2005/12/11 05:38:34 ↓ 05:37:06	両国2, 天気: 2005/12/12 20:21:09 ↓ 20:16:00

東大廣瀬・谷川研究室

東大相沢研究室



実現レベル

- Goo からだログ (NTTレゾナント)
- 体重、ウエスト等を測定し、日々記録する
- FoodLog (相澤研究室)
 - 日々食べた食事を携帯で撮影
 - 自動的にバランス等を算出



The image shows two screenshots of web applications. The top screenshot is the 'Goo Karada Log' interface, displaying a user profile for 'tadamaaan' with a weight graph for the period from August 15, 2009, to September 13, 2009. The bottom screenshot is the 'FoodLog' interface, showing a calendar for January 2009. A blue circle highlights a date on the calendar, with a blue arrow labeled 'クリック' (click) pointing to it. Below the calendar, a detailed view of a food entry is shown, including a photo of a bowl of food and a table of nutritional analysis. The table lists categories like grain, meat, bean, vegetable, fruit, dairy, drink, sweets, and rating, with corresponding values and sliders. A blue arrow points to the table with the text '結果が間違っているものは訂正してください' (Please correct the results that are wrong).



ライフログを支える技術

- ウェアラブルコンピューティング/ユビキタスコンピューティング
 - 携帯電話/GPS
 - HMD (Head Mount Display)
 - (Wii、タニタの体重計)

無意識下で情報を獲得したい (発信したい)



情報発信の変遷

- 特定の場所で、特定の人が
 - マスコミュニケーション
- 個人ホームページ/ブログ
- Twitter/Google Latitude (場所情報)
 - 「つぶやく」 (視聴者をイメージしないが発信する)
 - 携帯デバイスとの高い親和性、各種デバイスとも高い親和性
 - おならをつぶやく椅子
- ライフログ





情報にさらされる医療

- 様々なロギング/情報発信手法が実現
 - ブログ、ライフログ
 - 病院情報システム→医療行為のロギング
 - 映像、音声、位置情報、モーションキャプチャ
- 医療が求めるか否かは別問題

「医療」はどんな情報が必要で、どんな情報が必要でないか峻別することができるのか？



肥大化するログ

- 医学/医療が要求するログを獲得
 - 病院情報システムの肥大化
 - なんかデータ使えそう→マイニング
 - ↑現在の医療情報学はこのあたりがターゲット
- ユビキタス/ネットワーク社会
 - ブログからライフログへ
 - 監視カメラ/Nシステム/クレジットカード
 - 意識的・無意識的を問わず



「たいせつな」権利

- プライバシー権

- 私生活の自由：自分の情報は自由に使わせない

↑現在の医療情報学はここが議論の対象

- 自己情報のコントロール権

- Monster Patient化する患者さん

↑むしろこちらの方が医療に大きな影響を与えるのでは？

想定されるシナリオ

- 医療者に患者の情報が「PUSH」される
 - 投薬した薬剤を服用した/しない
 - 他の病院での検査結果が送信される
- 患者が全ての情報を「開示」している (=医療者が患者の情報を「PULL」できる)
 - 実際医療者はデータを閲覧できる
 - 開示された情報を見なかった場合、責任は誰に？





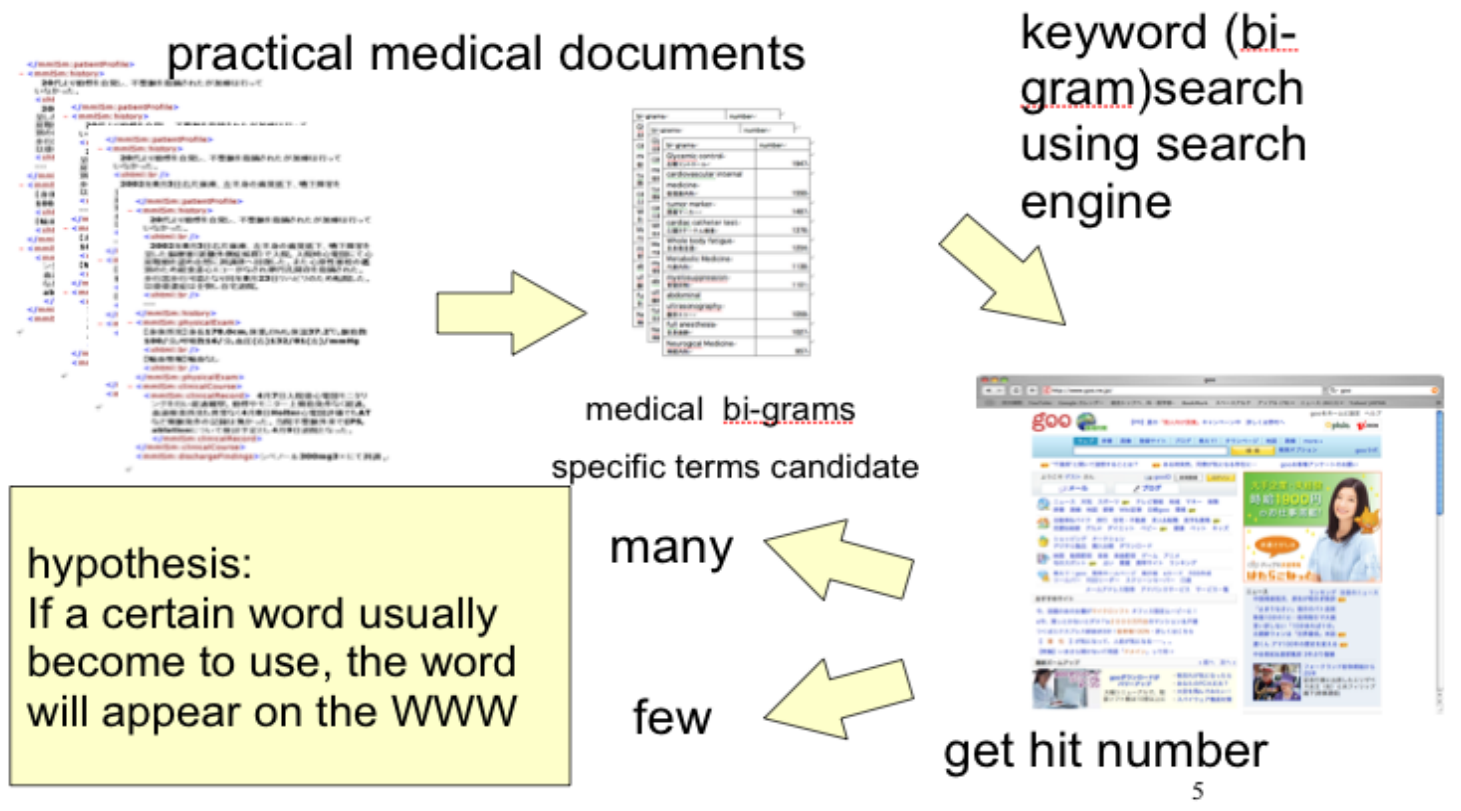
とりあえず、どんな情報に曝されるのか

- ブログ／webコンテンツ
 - 人々の情報の発露 (=ログ)
 - どんな知識が出現しているのか？
 - 医療の話題
 - 病院・医師の評判 (うわさ)
 - 京大病院情報システムのうわさ、メールシステムのうわさ
 - 治療法 (健康法)



用語性判別 (Googleに聞いてみよう)

今、どんな用語が使われているの？



医療web検索エンジン

- MANDARA (まんだら)
- 医療専門職か否かで検索結果を返る
- 検索質問：キーワード＋カテゴリ
 - キーワードマッチング
 - カテゴリによる重み付け
 - 静的スコア
 - TF・IDF値



まんだらや (城崎温泉)



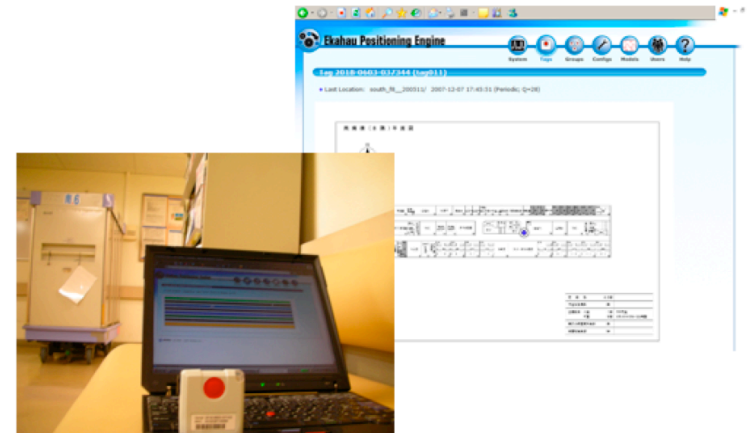
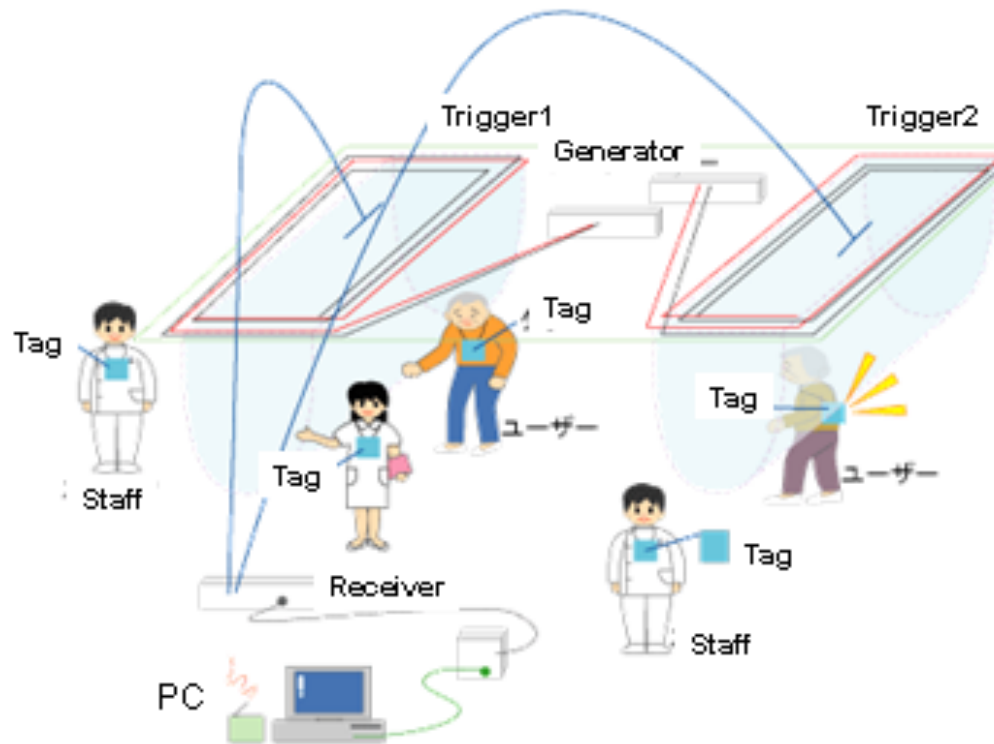
インタフェース

The screenshot displays the MANDARA Medical Web Search interface. It features a search bar at the top with a '検索' (Search) button. Below the search bar, there are two main sections: '一般' (General) and '医療従事者' (Medical Professionals). The '一般' section includes a list of medical specialties such as 移植外科学 (Transplant Surgery), 眼科学 (Ophthalmology), 救急医学 (Emergency Medicine), 外科学 (Surgery), 形成・再建科学 (Reconstructive Surgery), 血液内科学 (Hematology), 膠原病科学 (Rheumatology), 呼吸器外科学 (Thoracic Surgery), 呼吸器内科学 (Pulmonology), 歯科口腔外科学 (Dentistry), and 産科婦人科学 (Obstetrics and Gynecology). The '医療従事者' section includes 耳鼻咽喉科学 (Otorhinolaryngology), 循環器内科学 (Cardiology), 小児科学 (Pediatrics), 小児外科学 (Pediatric Surgery), 消化器外科学 (Gastroenterology), 消化器内科学 (Gastroenterology), 神経科精神科学 (Neurology and Psychiatry), 神経内科学 (Neurology), 心臓血管外科学 (Cardiovascular Surgery), 腎臓内科学 (Nephrology), and 整形外科 (Plastic Surgery). The larger version on the right also includes 放射線科学 (Radiology) and 麻酔科学 (Anesthesiology). A '検索' button is located at the bottom left of the interface.

石田、竹村, 2006

– 医療従事者／非医療従事者2つのインタフェース

位置情報ログ



Ekahau Positioning Engine
Aeroscort/Cisco

ME機器の利用情報をHISに飛ばす

Takemura, Kuroda, Kume, 2008

徘徊老人/介護従事者の位置情報取得
→自動日誌作成

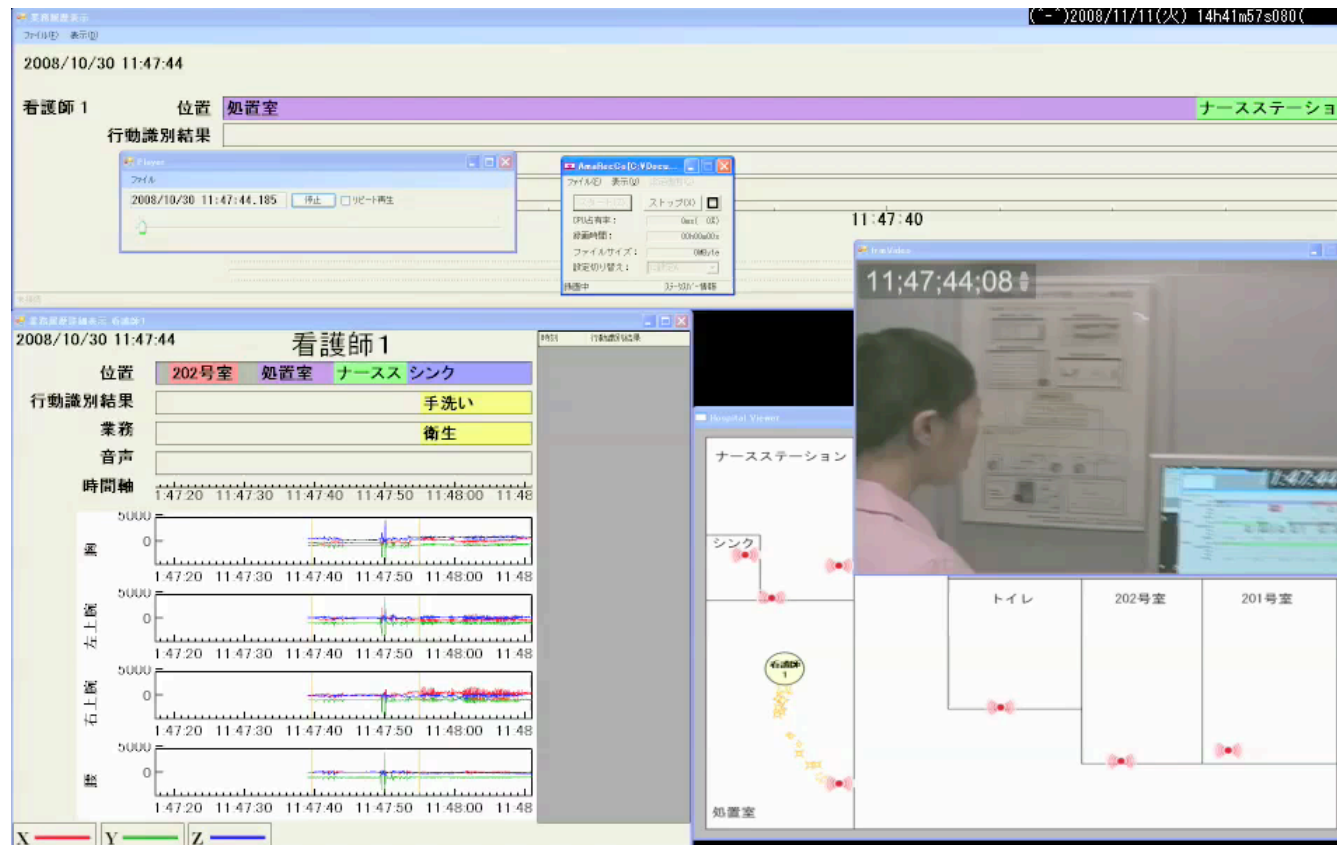
Ashida, Takemura, 2007

→ウェアラブルでバイスを組み合わせて
位置と動きで情報を同定



ライフログ×医療

e-ナイチンゲールプロジェクト



目的が所与的なら、作り込みが可能（指示受けを行えないか？）



言語/ライフログ/医療

- 医療における言語情報（電子カルテ）は、医療が情報にさらされる序章
- 目的が所与的でない「マイニング」という研究スタンスが生まれる
- 結果も「医療が求める」精度かは分からない
- しかし、背後に「ライフログ」という情報が控えている
 - 挙げ句、送りつけられる可能性がある
 - その時、医療はどう対応するのか？



まとめ

- 言語もデータとして扱える環境が整いつつある
- 言語情報に限らず、「医療」が要求および再利用を意図しない情報にさらされつつある
 - マルチメディア、位置情報、モーション、ログ
- 実学としての医療が受け止める準備があるかはわからない
- 知識のひとつの発露形態としての言語情報およびよりプリミティブな情報であるライフログから、今後の医療情報（学）のあり方を模索していきたい



最後に

医療の情報化から

→情報化する医療へ