

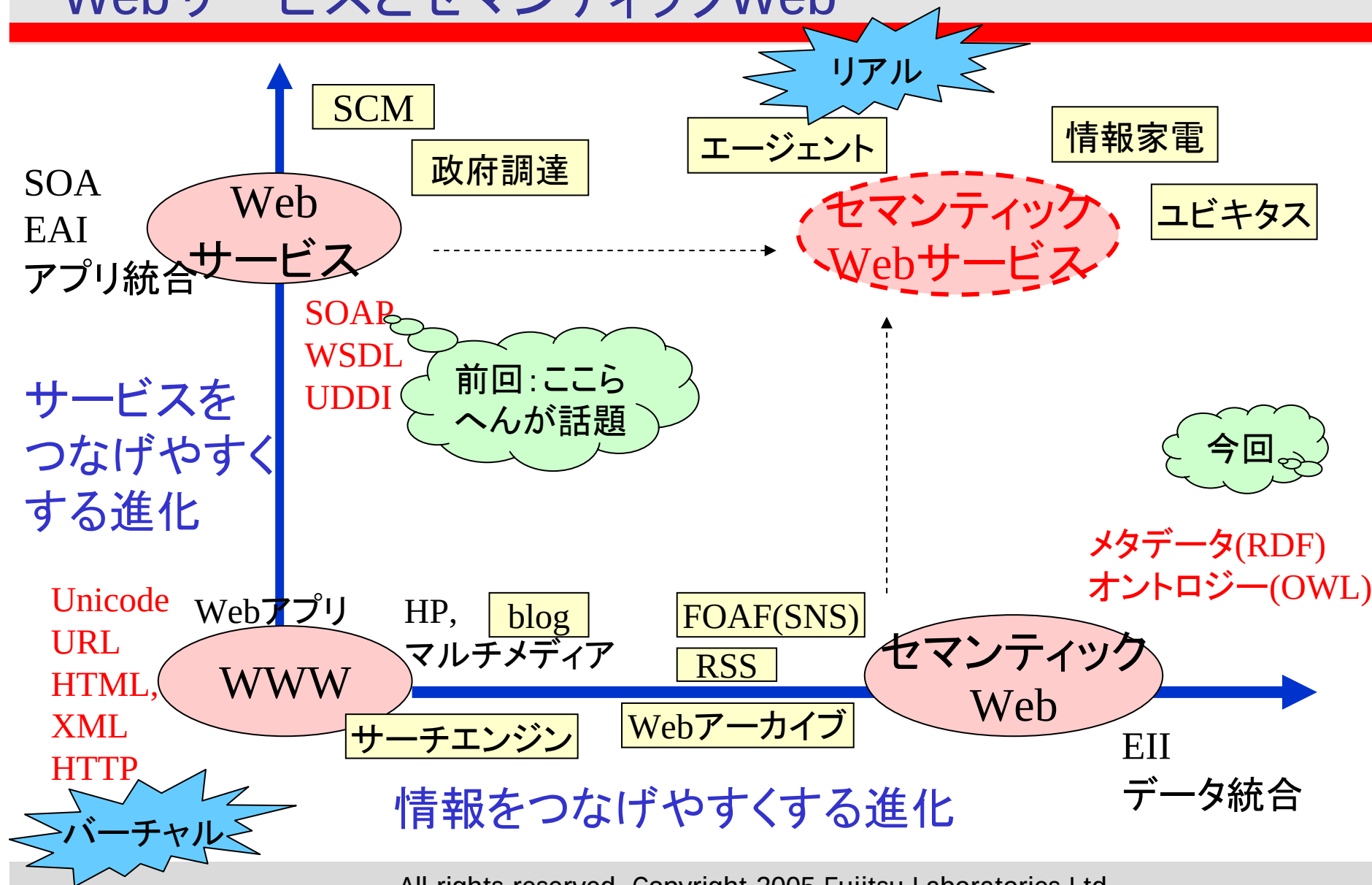
セマンティックWebとその展開 ～次世代の情報・サービス検索・活用～

2005.4.12 総務省情報フロンティア研究会
株式会社富士通研究所
知能システム研究部
津田 宏

htsuda@jp.fujitsu.com

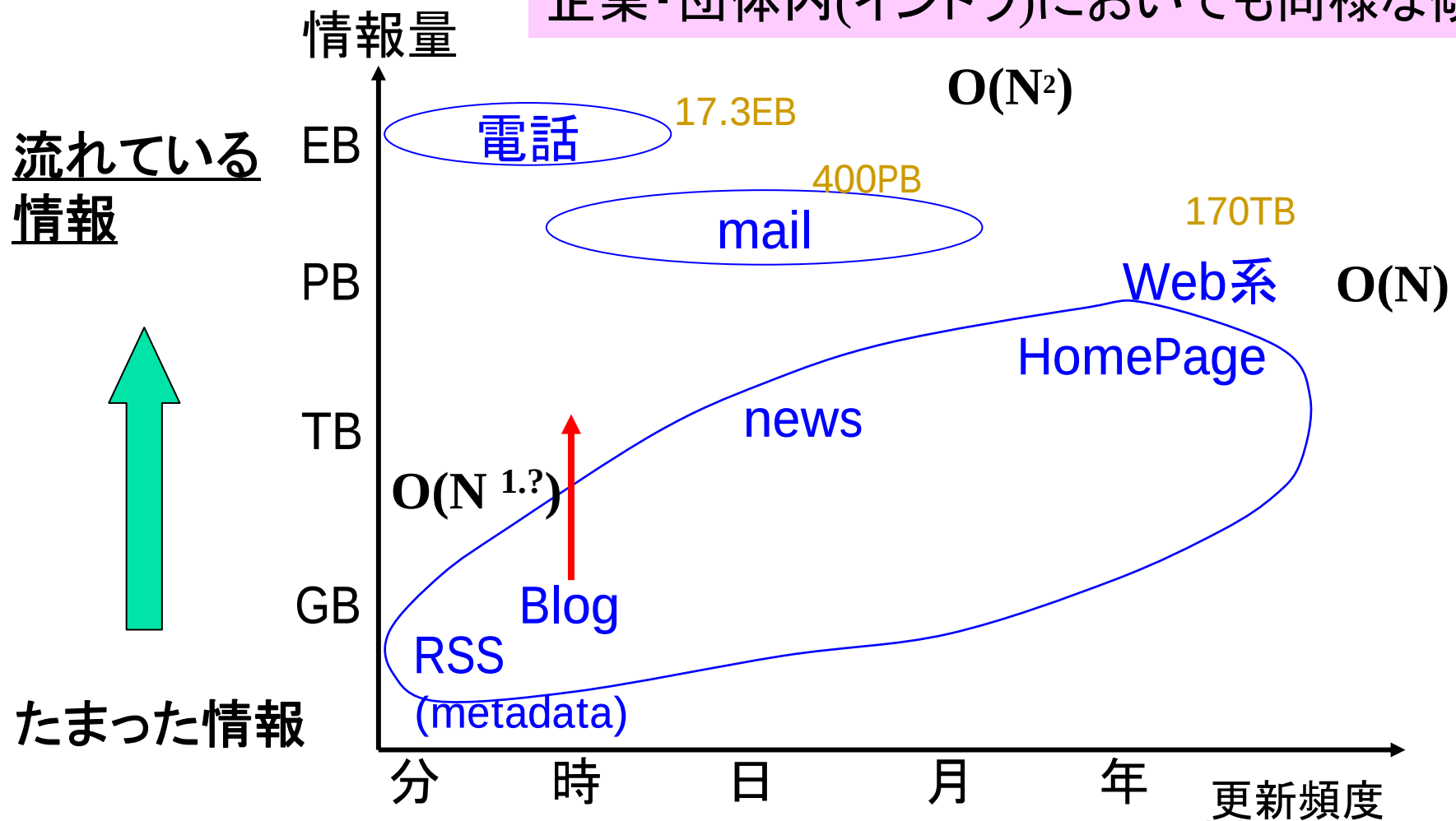
次世代Webの2つの軸:

WebサービスとセマンティックWeb



Webにおける情報量の爆発

企業・団体内(イントラ)においても同様な傾向



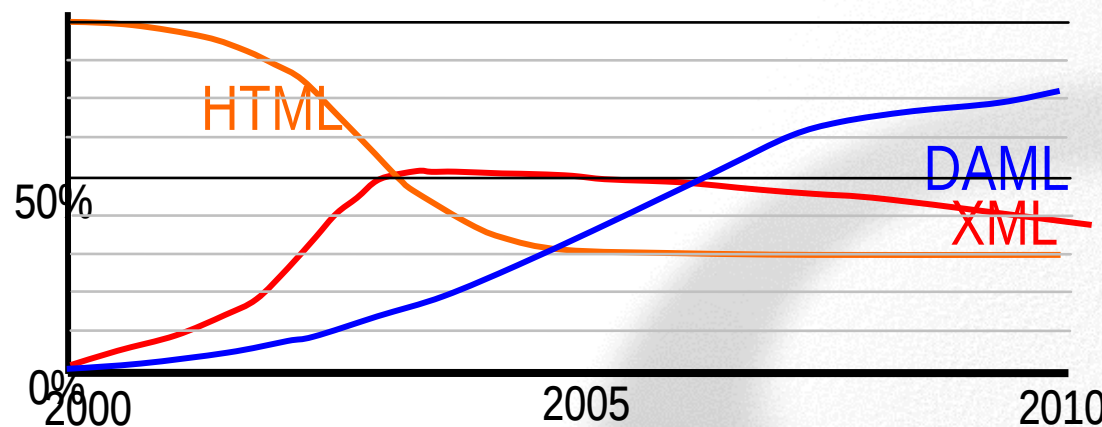
How much information 2003? より

Webのこれから(予想)

■ 情報過多の背景

- Webページ量の激増: 結果も膨大、人手で見る?
- Web中のゴミも増加: 信頼度まちまち、人手で判断?
- 検索対象の多様化: text, 画像, 動画, 人, サービス, ...

■ 今後: XML, メタデータ比率は増加の予想



きちんとした
情報を流そう
という動き

(by Mury Burke,
SWMU2, 2001.11)

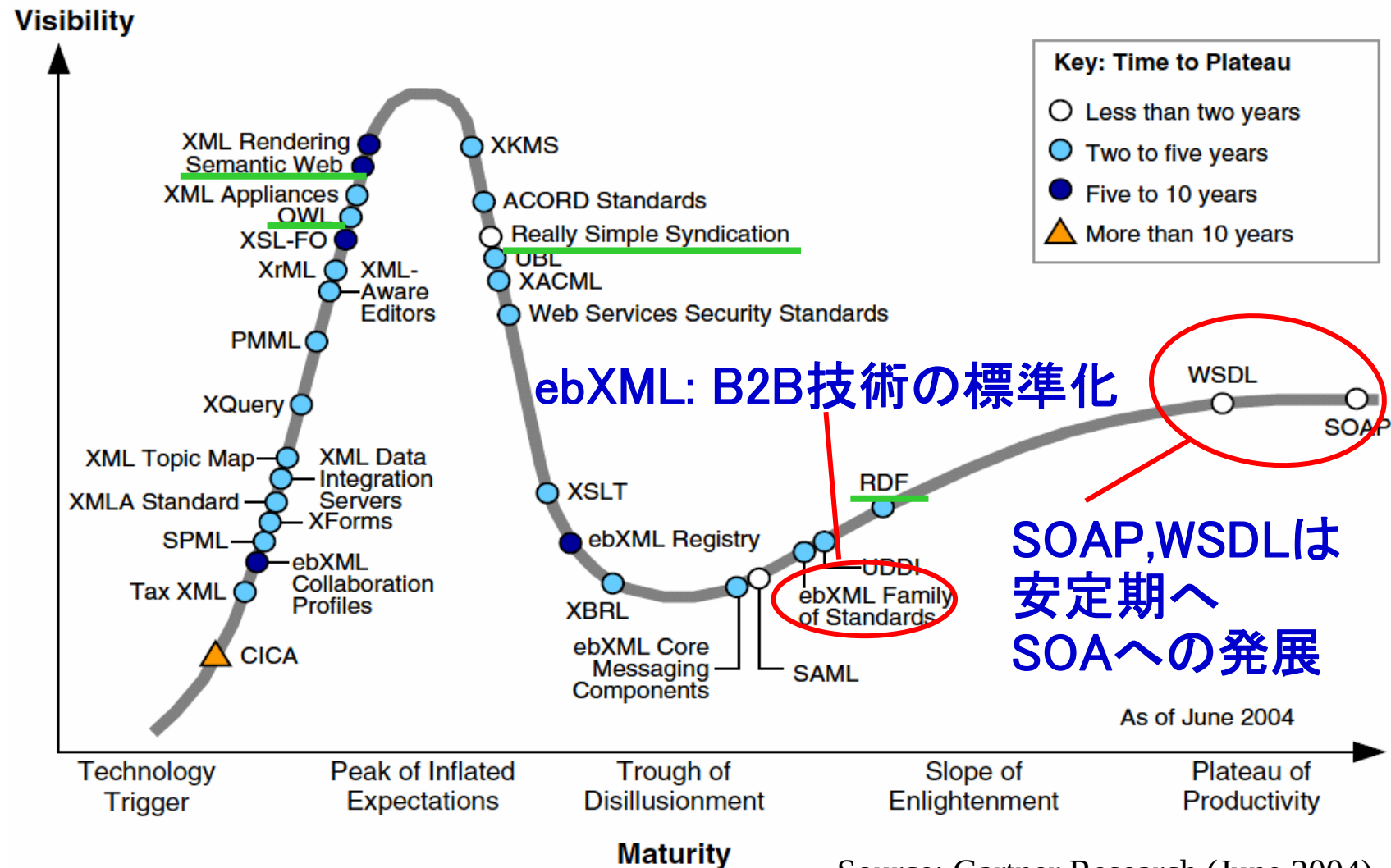
<http://www.daml.org/meetings/2001/11/swmu2/>

XMLに関する標準化のサイクル

(参考)

5

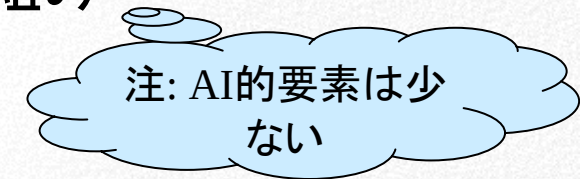
～Webサービスは、SOA、(企業内)B2Bおよび情報家電やユビキタス・ロボットなど、適用分野が拡大～



Source: Gartner Research (June 2004)

Semantic Webとは

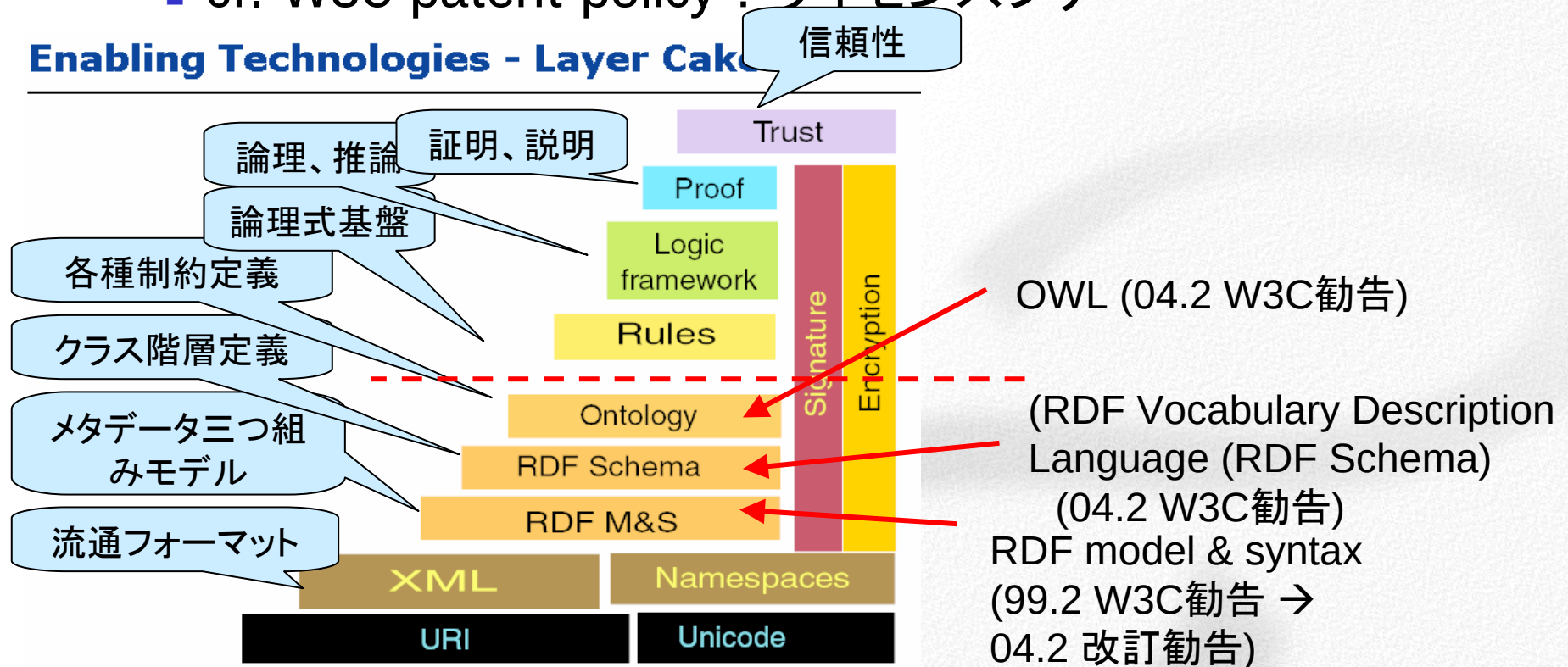
- Webの発明者であるT.Berners-Leeが提唱。Web上の情報を機械的に処理(検索、活用)するための枠組み
- キー技術:
 - **メタデータ(リソースに対する付加情報)**
 - 「蕎麦屋」ページに対する、「店名」「住所」「電話番号」のような付加情報。ページは人が見るもの。メタデータは機械処理するためのもの。
 - **オントロジー(メタデータを記述する語彙の定義)**
 - 「蕎麦屋」「寿司屋」などもののクラス、「店名」などの属性名に使われる用語関係を規定
- 実用化に向けたプロジェクトが各国で開始されている
 - W3C (World Wide Web Consortium) : Web系規約の標準化団体
 - 米: DARPA DAMLプロジェクト(2000-04)
 - 欧: EU-ISTプログラム(On-To-Web, SWWS, etc.), 電子政府
 - 日本: 国レベルの取組みはないが、注目度は高い
 - INTAPセマンティックWebコンファレンスは毎年開催、盛況



セマンティックWebの規格階層

- W3Cにて、メタデータ、ロジックの一連の記述階層と記述手段(入れ物)を標準化
 - cf. W3C patent policy : ライセンスフリー

Enabling Technologies - Layer Cake



(WWW2002, W3C trackより)

All rights reserved, Copyright 2005 Fujitsu Laboratories Ltd.

Webサービスと情報交換に必要な技術 (仕様) (参考)

～着実に揃いつつあるが、仕様の利用条件が課題～

		ebXML		RosettaNet	
ビジネス プロセス	ビジネスプロセス (コンテンツ)	UBL ★	STAR ★	JEITA ★	
	ビジネスプロセス (記述方法)	BPEL4WS ★			ebXML BPSS ★
標準 伝	コード(コンテンツ)	JANコード ★	DUNS Number ★	EPC ★	UCC ★
	要素項目集 (コンテンツ)	UBL ★	JADM ★	OAGIS CC Library ★	UN/CEFACT CC Library ★
	XMLスキーマ生成規約	UBL NDR ★	OAGIS NDR ★	UN/CEFACT ATG2 NDR ★	
	業務伝票の作成手法	UN/CEFACT Core Component Technical Specification / ebXML CC ★			
	伝票(フォーマット)	UBL ★	JADM BOD ★	STAR BOD ★	EDIFACT ★
			BOD(OAGIS) ★	JEITA ★	CII ★
プラットフォーム プロトコル	システム/リソース管理	WS-DM	WS-RF ★		
	トランザクション	BTP ★	WS-CAF ★		
	ネゴシエーション	WS-Policy			ebXML CPPA ★
	メッセージング/イベント	WS-Reliability ★	WS-Addressing ★	WS-RM	WS-Notification ★
	セキュリティ	WS-Security ★	Liberty ★	WS-Trust	WS-Federation
	レジストリ	UDDI ★	WS-Inspection		ebXML RR ★
	ベーシックプロトコル	SOAP ★	SWA ★	HTTP ★	WSDL ★
		JMS ★	HTTPS ★	IIOP ★	MIME ★
					BP AP SSB ★
		標準化済	標準化中	公開のみ	★ IPRフリー ★ IPRの扱い検討中
					WS-I

なぜWebで情報を探せないか？

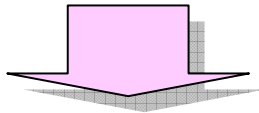
「今日開いている品川近くの歯医者はある？」

[現状のWeb検索] = “bag of words” (ページは単語の集まり)

「品川 AND 歯医者」で結果を一つ一つ人が調べていく

→ ゴミ: 日記, 千葉にある品川医院 ...

モレ: 高輪デンタルクリニック



[Semantic Webでの検索]

- ページに**メタデータ**をつけて、ゴミ・もれを減らす
- 検索結果を機械的に(ソフトウェアが)処理
(eg) 全部の結果を1枚の表にまとめる、
医者を予約しスケジュールに入れる



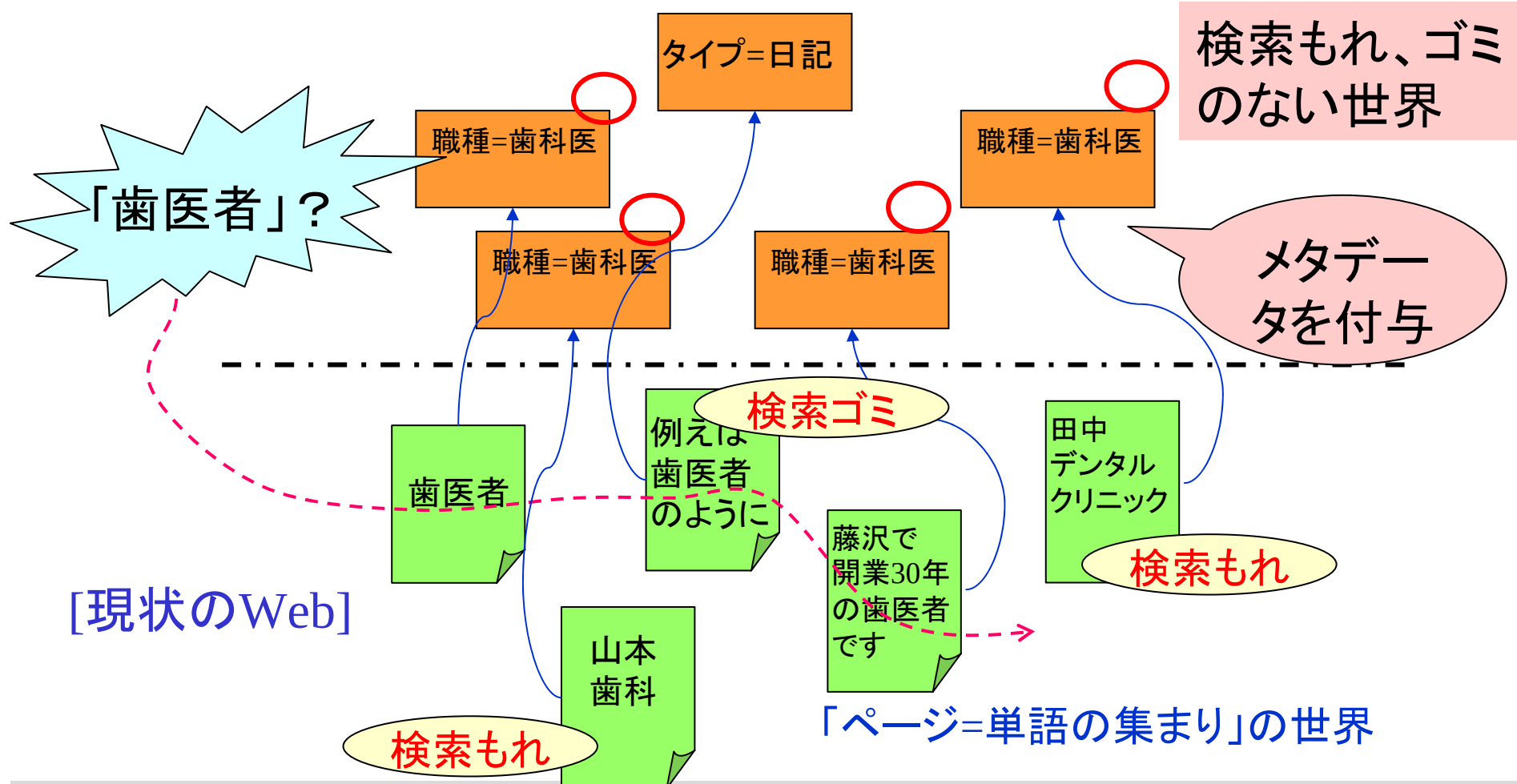
[メタデータ]
住所、開店時刻

...

検索例(1)「歯医者」のページ

[Semantic Web]

メタデータ(意味)の世界



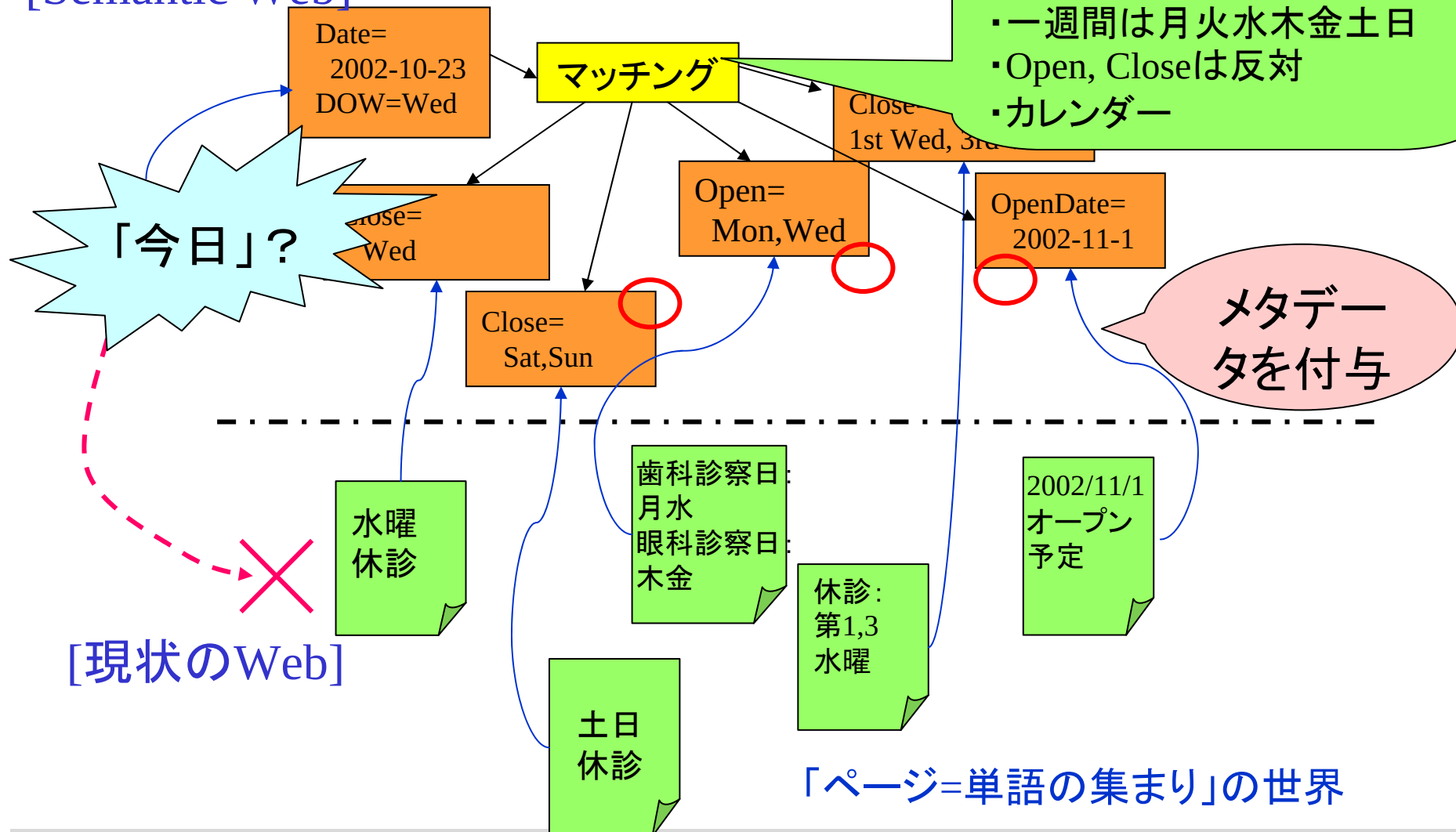
検索例(2)「今日開いている」

状況に依存
した検索

[Semantic Web]

【背景知識】

- ・一週間は月火水木金土日
- ・Open, Closeは反対
- ・カレンダー



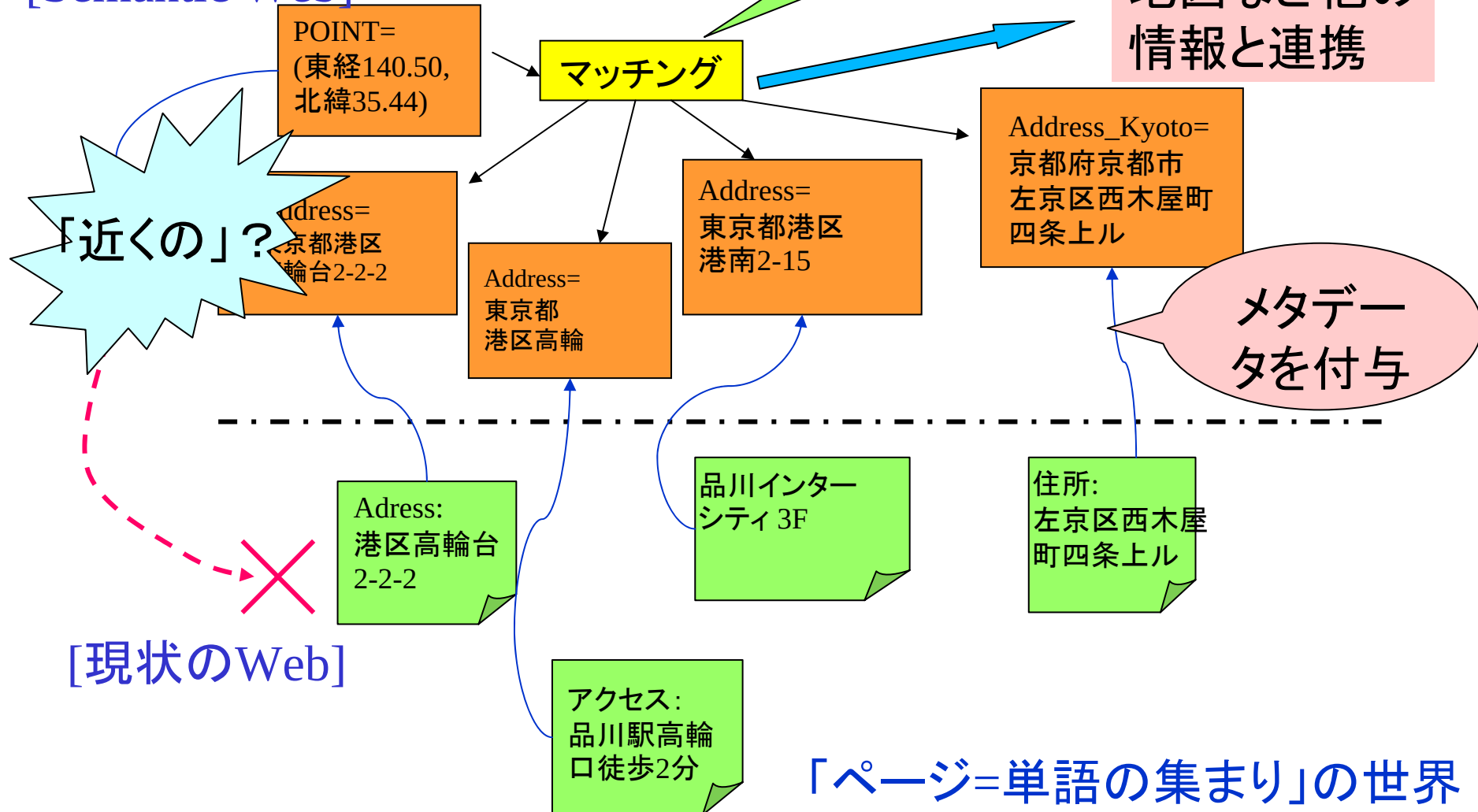
検索例(3)「近くの

【背景知識】

・地図(緯度経度変換)

[Semantic Web]

地図など他の
情報と連携

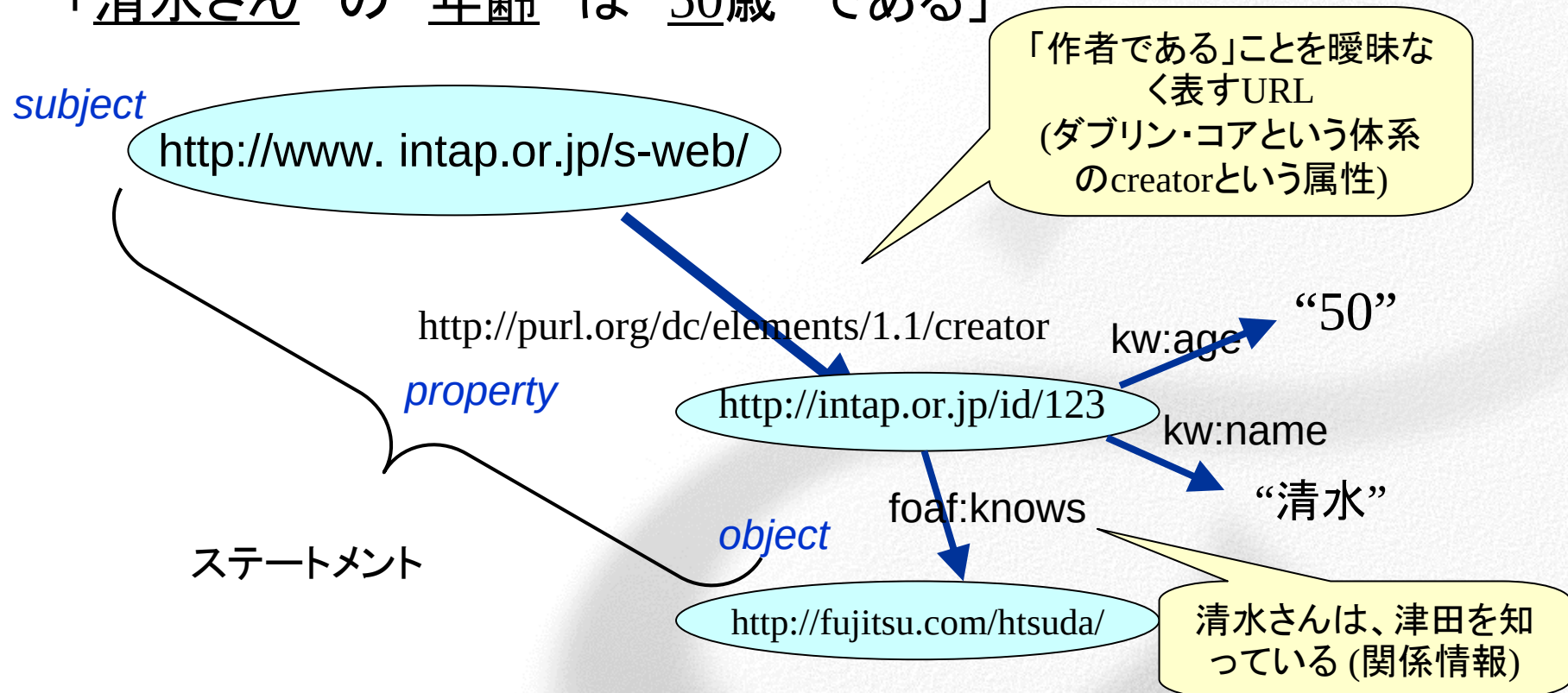


RDF (Resource Description Framework)

- (Subject, Property, Object) の三つ組みでメタデータを表現するフォーマットを規定

「http://www.intap.or.jp/s-web/ の 作者 は 清水さん である」

「清水さん の 年齢 は 50歳 である」



オントロジー(schema)

- オントロジー：対象とするドメインの基本となる「もの」と属性(関係)を記述する語彙を定義
 - (eg) Yahoo!におけるカテゴリ体系、MeSH(Medical Subject Headings)における生物医学専門用語階層、DublinCoreにおける書籍の15属性、
- メタデータがあっても、別々の語彙を使っているのは意味がない。そこで、
 - 語彙や制約を定義する言語を共通化: RDFS, OWL
 - なるべく共通語彙を使おう: eg) dc:creator
 - と言っても、世界共通は難しいのでいざとなれば語彙を変換してメタデータを相互活用

セマンティックWebサービス

- セマンティックWeb + Webサービス
- 現Webサービス: サービスの検索、合成に課題
 - つなげることに主眼。あらかじめ場所・呼び出しがわかっているものをつなげるケースも多い
 - UDDI: サービスの入出力データ構造検索はできるが、意味的検索ができない。現実には人が判断
- セマンティックWebサービス
 - Webサービスのプロフィール記述にメタデータやオントロジーなどセマンティックWebの技術を活用
 - OWL-S (W3C)
 - サービスプロフィール: このサービスがなにをするものか
 - サービスグラウンディング: サービスへのアクセス方法
 - サービスモデル: サービスがどう動くか

(参考) Webサービスにおける語彙統一

UN/CEFACT Core Component

16

- 業務伝票の作成手法に関する仕様
 - ISO15000-5 (ebXML標準) : UN/CEFACT Core Component Technical Specification (CCTS)
 - ebXML Core Components仕様をUN/CEFACTで標準化し、ISOでebXML仕様の第5部(ISO15000-5)として承認
 - 業務情報をモデル化する技法
 - XMLやEDIFACTといった構文からは独立
 - 中心概念は以下の2つ:
 - 業務に依存しない「コア構成要素」(CC)
例:「金額」
 - CCに業務上の文脈を付加した「ビジネス情報実体」(BIE)
例:「請求金額」
- 要素項目集(業務伝票に使う項目名のライブラリ)
 - 汎用的かつ抽象的なもの(例:住所)
 - UN/CEFACT Core Component Library (CC Library)
 - 特定目的かつ具体的なもの(例:発注者住所、納品先住所)
 - 各業界団体で、上記の汎用的な項目名をベースに作成する。

(例1) RSSによるblog/ニュースの新着情報閲覧

- RSS (RDF Site Summary, Rich Site Summary) : ページや記事のメタデータ

RSSリーダ

blog



新着情報を
統合閲覧

ニュースサイト

【従来】
人手でサイトに
アクセスして、
新着をチェック

RSSリーダが
各サイトの新
着情報を自動
チェック

[課題] ・RSS規格の混乱

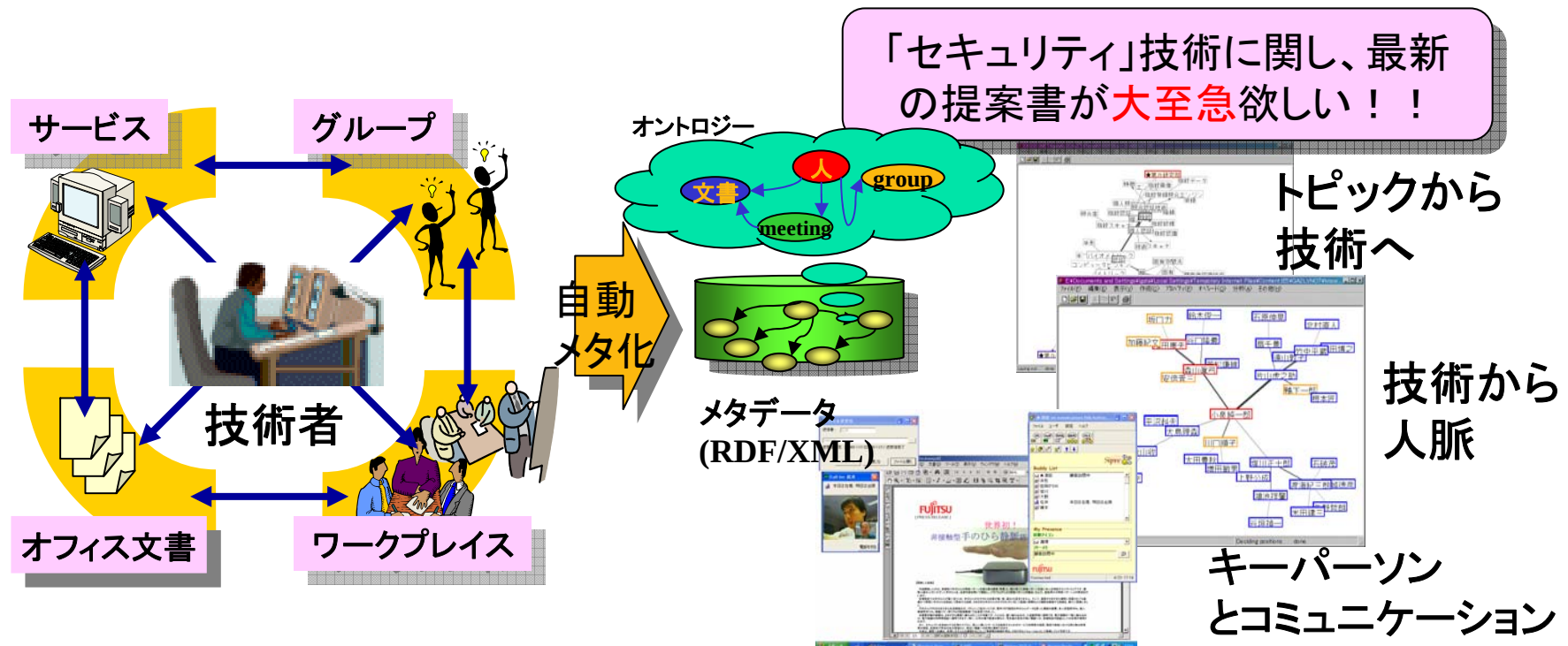
・ビジネスモデルで”push”を乗り越えられるかが普及の鍵？

(例2) 次世代検索としての人の検索(KnowWho)

- 人や人脈(コミュニティ):今でも重要だが従来型検索では探すのが大変
 - 「業務の情報の50~75%は人から直接得る」(Gartner 2002)
 - 人にメタデータを自動付与することで対応

- (eg) ヒューマンナレッジナビゲーター

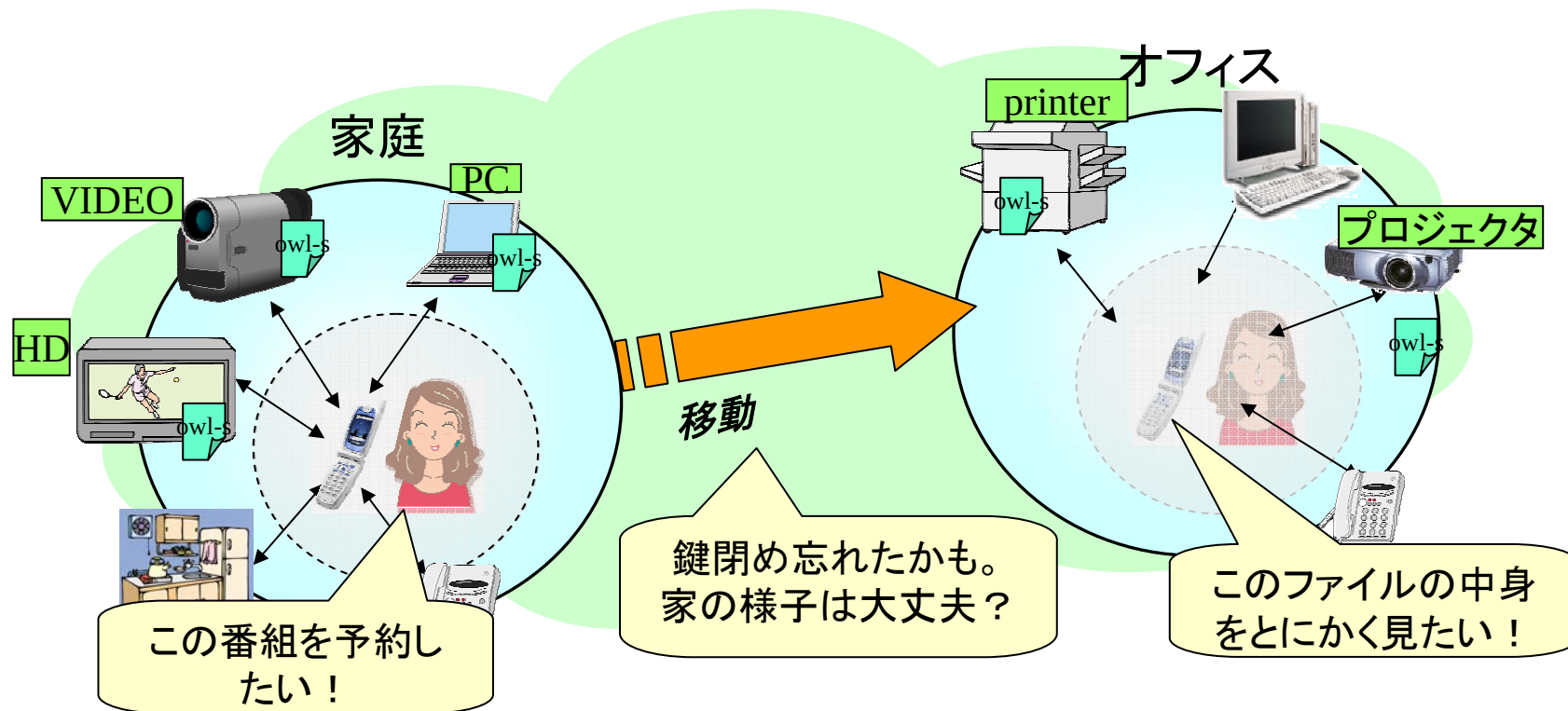
[demo]



目標: ワークスタイルを変革する検索

(例3) 情報機器の検索・連携

- 情報家電、機器：将来的にはWebサービスの対象に
 - 情報機器にメタデータがつき、場面に応じて使える機器を柔軟に探し・組み合わせ
 - (eg) 東芝 サービスマッチメイキング、富士通研 Task Computing
 - 課題：語彙(オントロジー)の標準化

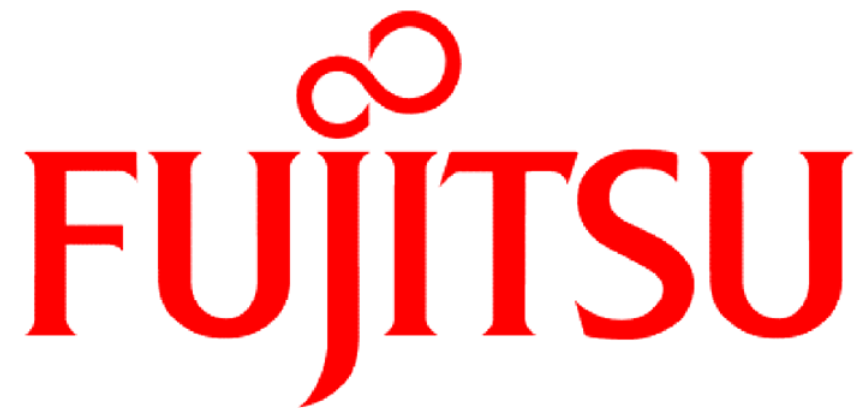


課題：誰がメタデータを書くか？

- メタデータデッドロック
 - 「メタデータがあればこんな良いサービスが作れるのに…」VS「こんなサービスがあるのだったら、ちょっと大変だけどメタデータを作っても良いのに…」
 - 費用対効果が見えない。(HTMLとの相違。ontologyも同様)
- メタデータを簡単に作る技術が望まれている
 - blogツール: コンテンツと同時にメタデータ(RSS)を出力
 - メタデータ自動抽出: 自然言語処理などのAI技術を利用し、コンテンツから半自動抽出
 - これらによって、早期サービス立ち上げ→メタデータが更にリッチに→さらにサービス拡充 ... というポジティブフィードバック！
 - メタデータ公開によるモチベーションも重要

課題：セマンティックWebにおけるトラスト

- Web: 性悪説、信頼性がまちまち
 - (eg) サーチエンジンに対するワードスパム攻撃
 - 良く検索される語をMETAタグに大量に入れることで、サーチエンジンを騙してページ露出度を上げる誤ったSEO(サーチエンジン最適化)
 - この後、METAタグは有効活用されていない
 - 階層最上位の「Trust」が伴わないとだめ。
 - 検索ならまだ良いが、サービス連携でスパムが入ったら？
- トラストや「安心できるWeb」への社会的な仕組みも合わせた対策が急務
 - 信用の認証局 vs. 大量の人、リソースを基本にした信用
 - Social Networking, (eg) FOAF, Friendster, LinkedIn



THE POSSIBILITIES ARE INFINITE