行政院國家科學委員會專題研究計畫成果報告

國科會 Web Technology 規劃推動小組

計劃編號：NSC 90-2213-E-002-147

執行期限：90/08/01 ~ 91/07/31

計劃主持人：項潔 教授

國立台灣大學資訊工程學系

目次

壹、前言.....p. 1

貳、Web Technology 規劃重點.....p. 2

參、Web Technology 推動情形.....p. 8

肆、研究成果資料.....p. 9

伍、未來展望、自評與建議.....p.12

陸、結語.....p.14

柒、附錄

九十一年度『Web Technology 研究』計畫 徵求通告.....p.14

國科會 Web Technology 系列研討會(一)講義

國科會 Web Technology 系列研討會(二)講義

壹、前言

1993 年全球資訊網（World Wide Web）的興起宣告全球數位化時代的來臨。短短數年間，電腦與網路已從少數人的研究工具變成多數都會人的生活必需品。 Web 所帶來的資訊產生與傳播的快速使得政府與產業均面臨前所未有的挑戰，一個政府或產業的競爭力往往決定於資訊取得的質量與速度。

除了資訊革命的大趨勢外，我國更面臨製造業外移與產業空洞化的危機，急需落實知識經濟的實質方案。而發展適合我國整體需求的 Web Technology，是決定我國在數位競賽中是否脫穎而出的重要一環。國科會工程處特別邀請產學研專家組成推動小組，針對 E-Government 、 E-Education 、 E-Market Place 、 E-Business 、 E-Entertainment 等 Web 應用上所需之技術，提出『Web Technologies 研究規劃』，列舉研究重點，以期能帶動學術界積極投入 Web Technologies 關鍵技術及應用之研究。鑑於此迫切需求，我們建議成立「Web Technology 規劃小組」，希望在 web 應用有需要的技術研究上，做一整體的規劃。

在未來的全球經濟體的架構下，國界不再是金流與物流的最大限制。Web 提供跨產業垂直整合的環境，使得供應面、製造面、需求面與金融面能在同一介面下直接互通，一體成型，不受空間的限制，這是未來 e–商業的模式。要達到這個理想，尚有努力的空間，本規劃案將探討 e–商業所需的技術，從垂直整合的角度作全面的規劃，以期台灣在 web 應用的技術上能培育出足夠的研究能量、成果與人才。

政府的 e–化是另一各重要課題。政府與企業 e–化的內涵與目的不盡相同，所以所需的技術與研究也有所差異，我們也希望在此推動計劃中確認所需之核心技術及推動研究。

從國家整體 e–化的發展來看，網路的架構（infrastructure）經過國科會「國家型電信」計劃及固網的建置，在硬體建設及研究上已獲得豐碩的成果。從內容面來看，國科會「數位典藏」國家型計劃、文建會「國家文化資料庫」計劃、研考會對政府資訊電子化的規劃及各大企業本身電子化的努力，在數位化上已累積了相當多的數位資料與經驗。Web technology 提供內容與硬體間所必須的軟體核心技術，使政府在 web 的內容建置、應用建置、軟體建置及網路硬體建置上達

到整體規劃的目標。

貳、 Web Technology 規劃重點

Web Technology 之分類與界定：

硬體建設類 (網路通訊設備)：網路硬體、網際網路骨幹、ISP 網路服務供應商、個人電腦與伺服器、...

Client / Server：

Client：	Hardware：PC, PDA, cell phone, etc.
	software：Browser, communication software.
Server：	3-tier, n-tier, firewall, web, database, application, etc.

軟體技術類：

A. 標準制定類：www.w3c.org (XML, RDF, HTML, etc.)
B. 資料安全類：身分認證 (PGP)、資料編碼 (SSL, SET)、網路攻擊反制、資料呈現機制、智慧財產權 (資料、網頁、程式), etc.
C. 資料處理技術：signal processing, pattern recognition, OCR, data mining, NLP, VOD (Video On Demand), 3D- graphic, etc.
D. 資料交換標準：EDI (Electronic Data Interchange), metadata interchange (XML-based), etc. 須結合 security.
E. 資料檢索技術：CBR (image, audio, video), NLP (Natural Language Processing), etc.
F. 網頁瀏覽技術：XML (layout：XSL, data structure：XML-schema), Java (Applet, Servlet), 3D-browsing, etc.
G. 資料庫類：Oracle, DB2, MySQL, etc.

工作項目包括：

1. 舉辦產學交流會，了解業界需求並擬定推動方針及研究重點。
2. 規劃 web technology 在國科會應推動的研究項目。
3. 規劃 web technology 之 RFP，並配合 91 年度計劃提出。
4. 舉辦每季工作會議，所有在 web technology 規劃中之計劃均應出席，交換研究心得與經驗。
5. 收及相關標準的資料，推薦我國應發展的 web technology 有關之

標準。

6. 舉辦研討會及成果發表會，介紹 web technology 相關研究與成果。
7. 協助業界舉辦與 web technology 相關之研討會，介紹相關產品。
8. 與業界洽商捐款支持相關研究經費及成立 web technology 實驗室。

一、XML Web Services Technologies

web services 強調以 XML 及軟體元件技術為基礎，開發 self-describing 及 self-contained 的模組化應用，使之能與其它的 web services 結合及搭配，以創造出新穎而有高附加價值的產品、程序（process）及價值鍊（value chains）。web services 在近來已成為熱門的技術和產業課題，台灣自應重視此趨勢，發展相關的技術和應用，包括

1. basic technologies：protocols、service description、service interfacing、service quality-of-service、service flows and process、service discovery、service transaction、management and policy、service security and access control、etc.。
2. interoperability：middleware interoperability、syntactic interoperability、structural interoperability、semantic interoperability、etc.。
3. middleware and tools：service authoring、service publishing、service discovery、service brokering、service creation and assembly、service integration、service collaboration、legacy system wrapping、etc.。
4. mobile web services：web services over mobile communication platforms、location-based services、location acquisition、etc.。

二、Web Service Applications

在 web services platform 上，發展架構於 Internet 的 dynamic distributed applications，以提供用戶更方便有效的數位服務，包括

1. services with rich content：auctions、airline quotes、news feeds、credit check、travel schedules、publishing、etc.。
2. transactional web services for B2B or B2C：customer relationship management、supply chain management、direct marketing、

dynamic pricing、etc.。

3.business process externalization：web-based EDI、web marketing、extended supply chain、business linkages at workflow、etc.。

4.location –based web services。

5.professional service automation：opportunity management、engagement structuring、project management、resource management、knowledge management、etc.。

6.knowledge service automation：experience sharing、knowledge capture、knowledge accumulation、knowledge search、knowledge categorization、knowledge extraction、knowledge retrieval、intellectual property brokering、patent map、risk assessment、decision making support、etc.。

7.e-government。

8.web-based learning：web-based collaborative storytelling system、virtual kids city to explore identity and values、digital learning community、etc.。

三、Semantic Web

基於 XML 和 RDF 技術發展對電腦有意義的新形式 web contents 及 services，研究課題包括：

1. basic technologies：knowledge representation to express domain ontology；agents to collect web contents from diverse sources、process the information and exchange the results with other agents；knowledge evolution。
2. visualizing tools：to browse domain ontology。
3. authoring tools：to edit domain ontology。
4. converting tools：to convert different ontology representations。
5. consistency maintenance of ontology。
6. creation and updating of ontology。
7. collaborative ontology development。
8. value-added services by semantic web。

四、Sociable Web Computing

以 web services 技術實現更豐富的社交方式和內涵，使得人際溝通及社交活動，可以打破時空的限制、文化限制、語言限制等，研究

課題包括：

1. Cyber Clubs：以 web services 方式，動態組構俱樂部提供的活動服務和活動場所，而且俱樂部成員可以在俱樂部的虛擬活動空間中自由游走、交談社交或參加團體活動。
2. Virtual Fashion and Cultural Dispersion：以 web service platform 建構成一個新興的社群平台，使其文化傳播力或流行感染力，皆高於傳統社群。
3. Universal Chat Circles：提供 universal language translation、synchronous text and audio conversation、dynamic grouping of conversational clusters、visual archival chat logs 等機能，以創造一個可以跨越國界、跨越語言的 online discussions 理想環境。
4. Live Web：讓 web users 可看到其它 users 的 real-time activities，進而了解 web population 的 behavior patterns，以增強 web community 的互動性和臨場感。

五、Web Intelligence

利用 AI 及 advanced information technology 於 web services，以建構 web-based intelligent information systems，研究課題包括：

1. web mining。
2. web warehousing。
3. web farming。
4. web knowledge discovery。
5. web event tracking。
6. hypertext analysis and transformation。
7. learning user profiles。

六、Web-Based Applications

1. Business Intelligence
2. Computational Societies and Markets
3. Customer Relationship Management (CRM)
4. Electronic Commerce and Electronic Business
5. Electronic Library
6. Price Dynamics and Pricing Algorithms
7. Measuring and Analyzing Web Merchandising

8. Web-Based Decision Support Systems
9. Web-Based Distributed Information Systems
10. Web-Based EDI
11. Web-Based Learning Systems
12. Web Marketing
13. Web Publishing

七、Web Human-Media Engineering

1. Art of Web Page Design
2. Multimedia Information Representation
3. Multimedia Information Processing
4. Visualization of Web Information
5. Web-Based Human Computer Interface

八、Web Information Management

1. Data Quality Management
2. Internet and Web-Based Data Management
3. Multi-Dimensional Web Databases and OLAP
4. Multimedia Information Management
5. New Data Models for the Web
6. Object Oriented Web Information Management
7. Personalized Information Management
8. Semi-Structured Data Management
9. Use and Management of Metadata
10. Web Knowledge Management
11. Web Content Management
12. Web Page Automatic Generation and Updating
13. Web Security, Integrity, Privacy and Trust

九、Web Information Retrieval

1. Approximate Retrieval
2. Conceptual Information Extraction
3. Image Retrieval
4. Multi-Linguistic Information Retrieval
5. Multimedia Retrieval
6. New Retrieval Models

7. Ontology-Based Information Retrieval
8.Automatic Web Content Cataloging and Indexing

十、**Web Agents**

1. Dynamics of Information Sources
2. E-mail Filtering
3. Global Information Collecting
4. Information Filtering
5. Navigation Guides
6. Recommender Systems
7. Remembrance Agents
8. Reputation Mechanisms
9. Resource Intermediary and Coordination Mechanisms
10.Web-Based Cooperative Problem Solving

十一、**Web Mining and Farming**

1. Data Mining and Knowledge Discovery
2. Hypertext Analysis and Transformation
3. Learning User Profiles
4. Multimedia Data Mining
5. Regularities in Web Surfing and Internet Congestions
6. Text Mining
7. Web-Based Ontology Engineering
8. Web-Based Reverse Engineering
9. Web Farming
10. Web-Log Mining
11. Web Warehousing

十二、**Web Information System Environment and Foundations**

1. Competitive Dynamics of Web Sites
2. Emerging Web Technology
3. Network Community Formation and Support
4. New Web Information Description and Query Languages
5. The Semantic Web
6. Theories of Small World Web
7. Web Information System Development Tools

8. Web Protocols
9. XML standards for various application domains/business sectors

參、 Web Technology 推動情形

規劃具體推動方向：（參考資料：http://www.itis.org.tw/forum/content3/01if14e.htm）

1. 加強網路建設：經由基礎建設技術之提升與整合，建置安全可靠之電子服務環境，能吸引眾多企業和消費者參與應用。而基礎寬頻網路之建設項目包含：
   A. 骨幹網路寬頻化。
   B. 用戶迴路寬頻及光纖化。
   C. 擴增國際海纜頻寬。
   D. 開發寬頻無線迴路。
2. 健全法規組織：為因應網際網路上之電子化應用順利運作，政府應制定明確之相關法規及標準規範。
3. 加強研發電子認證機制、網路資訊安全機制。
4. 有效運用資訊與通信科技，形成台灣各種企業 e 化之網路價值獲利模式。
5. 由民間加強將知識轉變為實際產業市場之運用，而政府致力於建立發展環境與排除營運障礙，以建立創新與創業機制及推廣資訊科技與網路應用為動力，已達成『知識經濟發展方案』之政策目標
6. 培育人才與科技水準，解決人才不足及獎勵技術升級。
7. 輔導相關產業升級，轉移相關技術提昇競爭力。
8. 定期舉辦研討會及觀摩會，促進產官學界交流，交換研究心得與經驗。

台灣現有高級資訊核心架構以及軟體之高級研發人力，仍然以在學術界的人員佔絕大多數。在產業界尚不能支持大型研發中心，而學術界又以資訊學門為最大單一學門的情況下，使得學術界結合學術研發能量以因應產業界迫切需求，成為 Web Technology 責無旁貸的責任。

此計畫自去年九月開始，舉辦過許多活動，包括規劃小組會議、研討會以及計畫徵求通告，參加人數踴躍，參與研討會的人數每次都

高達百人以上，其中不乏資訊相關科系之研究人才，對於 Web Technology 的推動，在實質上有很大的幫助。

會議中再次說明，此次 web technologies 計畫起始係由 Microsoft 公司建議，由主委交辦規劃，強調的重點為 web services，尤其是以 e-business 及 e-government 兩方面為主。Web technologies 一直是很重要的課題，希望能藉該計畫整合國內廣大的資管人才。Web technologies 所研究的課題範圍應廣，不應只局限在微軟提出的 web services 之下，而是要選擇對臺灣日後資訊發展有助益的研究。所提出的課題包括 information retrieval/extraction，information engineering, e-commerce, data-mining, e-learning, XML/RDF/XML tools, agent technology, access control and security, e-government, live web, semantic web 等，各課題仍側重 web services 方面的研究。

肆、 研究成果資料

在日新月異的今日，Web 對提升本國競爭力與產業獲益能力成為不可獲缺的關鍵因素。基於此一原因，國科會 Web Technology 積極鼓勵學術研究以及產學合作，目前已經擁有豐碩的研究成果，培養了無數的專業人才以及提高各研究機構之國際學術地位。

一、 91 年，在 34 件申請案當中，Web Technology 計畫下所通過的件數有 17 件與經費共有 8,517,600 元，內容包括以元件為基礎之電子書包自動同步化之研究、圖碼合一、流程導向的開發方法與 Web Services 的整合、以 XML 為基礎之適性化顧客關係管理資訊系統架構之建立、自動化 Web 網頁分類演算法之研究、一個以 ontology 為基礎的 semantic web content RDF/XML 產生器、以及數個與 Web Technology 與資訊相關的總計畫與子計畫。

91 WEB Technology 計畫申請核定統計一覽表

大葉大學資訊工程系	以元件爲基礎之電子書包自動同步化之研究	邱紹豐
國立雲林科技大學資訊管理系暨研究所	圖碼合一、流程導向的開發方法與 Web Services 的整合(I)	董少桓
中華大學資訊工程學系	以 XML 爲基礎之適性化顧客關係管理資訊系統架構之建立(I)	游坤明
元智大學資訊工程學系(所)	自動化 Web 網頁分類演算法之研究	楊正仁
元智大學資訊管理系(所)	一個以 ontology 爲基礎的 semantic web content RDF/XML 產生器	陸承志
國立政治大學資訊科學系	總計畫(I)	劉吉軒
國立政治大學資訊科學系	子計畫一：數位內容加值服務與互通性之發展研究(I)	劉吉軒
國立政治大學資訊科學系	子計畫三：以閱聽人資訊爲基礎的 Web 新聞系統之設計與實作(I)	沈錳坤
國立政治大學資訊科學系	子計畫二：以 VoiceXML 爲基礎的個人化資訊廣播站(I)	廖文宏
國立成功大學電機工程學系（所）	總計畫(I)	蔡尚榮
國立成功大學電機工程學系（所）	子計畫一：具良好延展性支援 Web Service 之 XML 資料存取系統(I)	蔡尚榮
國立成功大學資訊工程學系（所）	子計畫二：一個具有追蹤與預測使用者位置功能之行動網絡服務(I)	鄭憲宗
國立成功大學資訊工程學系（所）	子計畫三:於 Web 環境提供利用舊有應用系統功能之行動網絡服務(I)	朱治平
國立成功大學資訊工程學系（所）	子計畫四：行動網絡環境下之資料探勘機制及個人化行爲預測引擎(I)	曾新穆
長榮管理學院資訊管理系	子計畫五：智慧型分類/行動遞送網絡服務研發(I)	李健興
國立政治大學資訊科學系	建立代理者溝通機制：語意表示法與交換問題之研究	劉昭麟
國立政治大學資訊科學系	利用電子憑證來驗證開放式代理者電子化服務環境的可信度	胡毓忠

二、 與 Microsoft 合作落實知識經濟軟體人才培育提案中，微軟所提之執行方針：

1. 國科會成立 Web Technologies 推動小組
2. 推動校園軟體人才培育訓練計劃
3. 結合育成中心加速產業技轉，落實產業競爭力提昇
4. 成立校園 XML Web Service Research LAB 研究中心

網路技術校園專才培育計劃的目的，培育三百位校園網路技術專才，使其有新一代網際網路之技能。藉由成立線上社群，提供網路技術新知，持續交流研習心得，成為一長期性、自給自足式的網路技術人才庫。並由台灣微軟提供授課所需之教材，以利其於校園內傳播網路技術。課程內容一共有三程不同之課程，如下：

- 課程一：利用 Microsoft Visual C#.NET 開發 XML Web Service。
- 課程二：如何利用 ASP.NET 開發新一代的網路應用程式。
- 課程三：新一代的程式設計語言 C#。

這三門課分別教授 Web Service，Web Application，及 Middleware development 相關技術及開發。

師資，由業界知名之教育訓練中心(如恆逸、華彩等)挑選其微軟認證之資深講師多人擔任。教材，以 Microsoft Official Curriculum (MOC) 微軟官方教材為本。

所使用到之 MOC 項目包括：

- 2524A: Developing XML Web Services Using Microsoft Visual C# .NET Beta 2
- 2063B: Introduction to ASP.NET
- 2124B: Introduction to C# Programming for the Microsoft .NET Platform

成果驗收，各課程結束時，以考試驗收成果，通過者頒發結業證書。結業證書應以"台灣微軟認證之網路技術校園講師" 為頭銜，並冠以所修習之課程專才。後續計劃有：

1. 免費軟體贊助

台灣微軟將提供十套 MSDN Academic Alliance (包括所有作業系統、開發工具、伺服器軟體)給十所大專院校。由於數量有限，應採用 Request for Proposal 之方式，由各校或系所提出申請案，再由微

軟依申請案內容選擇十所。而 Request for Proposal 的內容將建議舉辦校內之訓練課程、比賽，並以導入 Microsoft .NET 技術為第一優先考量。

2. 持續技術交流

台灣微軟將成立一個 .NET 社群，以為學員後續討論及互相支援的交流中心，以維持訓練後的學習動力。並邀請受訓學員參予每週或月的專欄撰寫、問題回答、並討論彼此的授課心得及教材。

3. 頭銜及發展舞台

台灣微軟在日後的校園巡迴講座中，將邀請學員參與演說。並請傑出之學員於重要雜誌或網站上發表專欄。對於教育推廣有重大貢獻的學員，將給予進階之頭銜以為鼓勵。

三、 研討會

在此計畫年度，舉辦了數次研討會，包括在台大舉辦的第一次 Web Technology 系列研討會、在成大舉辦的第二次系列研討會，以及在中研院舉辦的 Semantic Web 研討會，三次的參加人數都十分踴躍，第一次內容有 Text anf Log Mining on the Web、Semantic Web、Web Sevices Technology and Applications，以及電子化政府入口網與各機關系統整合之共通平台初探。第二次內容有從 G 看未來的行動理財服務、無線網路服務的未來展望、行動式網際網路服務平台，以及為行動服務而設計之共通行動終端平台的探討與演講。第三次內容有 Building Sharable Ontologies for Intelligent Agents Based on Semantic Web、An Introduction to Semantic Web Portal、Semantic Web and Agents，以及 Applications of Semantic Web。開會資料放於附檔。

伍、 未來展望、自評與建議

自從我國政府在 80 年代將資訊工業定為國家重點發展項目以來，我國的資訊工業在 90 年代以驚人速度成長，在 1994 年資訊總產值已達到 160 億美元，並持續以 2 位數字成長，使我國成為繼美、日、德等國之後的資訊大國。根據資策會 MIC 分析全球資訊工業趨勢、公元 2002 年，全球總產值將達 9000 億美元，如何加強我國資訊科技的提昇、以繼續在全球資訊產業上佔有優勢，是我國政府的當務之急。

在政府大力推動高科技產業的政策下，使得我國高科技產業實力倍增，在世界上具有舉足輕重的影響力，扶植出今日蓬勃的高科技產業。今天面對新經濟泡沫危機的時點上，高科技產業仍應繼續推動，是之成為穩定經濟發展的支柱。我國的資訊發展水準在 2010 年時的發展目標是就技術、產品研發、軟體設計開發等多項要能和歐美及日本的先進國家並駕齊驅，成為亞洲之頂尖，遠遠領先亞洲其他國家。

經濟產業之發展可以活絡經濟之發展，強化社會之基礎。所以如何更加積極發展資訊產業將會建立資訊、經濟和社會之間的良好互動關係，並可提昇國家在國際間的政經實力、改善人民生活品質、厚實產業發展水準及增強社會行政效能。

為了配合上述的目標，建議將智慧型網路服務列為未來發展重點，是以 Internet 為溝通媒介，來完成不同服務提供者服務的連結與運作，讓資訊服務得以形成一種類似格狀，多方參與之服務網路。因此，智慧型網路服務除了具備內容認知的能力，還能透過多網路的傳遞功能提供整合性的服務，讓使用者能以更好、更容易的方式存取各種資訊與功能，使所需的服務能夠隨時隨地取得。近年來，網路服務以受到主要 IT 大廠支持，均已推出網路服務價購及技術平台，例如，Microsoft.NET MyService, IBM Dynamic e-Business, HP e-Service, Oracle Dynamic Service, Sun ServiceOnDemand 等等。根據 Gartner Group 估計，2005 年 75%之 BB 關係將採用網路服務技術，該技術預期將帶來新世代 B2C、B2B、System 及 Device 相關服務發展。

我國目前之資訊發展已經活絡台灣經濟實力及社會基礎，如消除距離障礙、均衡國土區域發展、再造企業組織架構、加速國家經濟流通、降低弱勢族群經濟落差，並改善老若殘障生活品質、提昇家庭及社區生活便利、再造就業環境及勞僱關係、強化社會福利執行效率及發揮環境保護水準。

期望在資訊領域推廣上，能帶動台灣及國際生活育樂 e 化，使全民都可以透過網際網路及資訊產品，處理大部分的日常事務，達到新素實簡的生活品質。不停地提昇政府行政效能，將能使民眾能夠享受到更方便的各項政府服務，以致大幅提昇民眾對政府的施政滿意度。

陸、 結語

培育高科技資訊科技研發人才，厚植我國資訊產業實力。期望在下一世紀，我國能擠身全球資訊大國之列，因此，我國資訊產業必須在技術研發能力上有所突破，尤其於硬體關鍵零組件研發製造技術方面，以及高附加價值之軟體開發計數之提昇上，有待產、官、學、研各方面配合。

在學界之發展上，除應鼓勵自由型之尖端科技研發外，配合我國資訊產業的策略方向，在產學互動之原則下，以產引學、以學導產，是目前及未來重要之推動原則。

柒、 附錄

九十一年度『Web Technologies 研究』計畫徵求通告

九十一年度「Web Technologies 研究」計畫

徵求通告

國科會工程處

資訊學門

民國九十年十一月二十七日

一、背景說明

1993 年全球資訊網（World Wide Web）的興起宣告全球數位化時代的來臨。短短數年間，電腦與網路已從少數人的研究工具變成多數都會人的生活必需品。 web 所帶來的資訊產生與傳播的快速使得政府與產業均面臨前所未有的挑戰，一個政府或產業的競爭力往往決定於資訊取得的質量與速度。

除了資訊革命的大趨勢外，我國更面臨製造業外移與產業空洞化的危機，急需落實知識經濟的實質方案。而發展適合我國整體需求的 web technologies，是決定我國在數位競賽中是否脫穎而出的重要一環。國科會工程處有鑑於此迫切需求，特邀請產學研專家組成推動小組，針對 E-Government, E-Education, E-Market Place, E-Business, E-Entertainment 等 web 應用上所需之技術，提出『Web Technologies 研究規劃』，並列舉研究重點，以期能帶動學術界積極投入 web technologies 關鍵技術及應用之研究。國科會對符合規劃之研究計畫將盡力支援，敬請各校教授及研究機構研提研究計畫。

二、規劃研究重點

(一) XML Web Services Technologies

Web services 強調以 XML 及軟體元件技術為基礎，開發 self-describing 及 self-contained 的模組化應用，使之能與其它 web services 結合及搭配，以創造出新穎而有高附加價值的產品、程序及價值鍊。建議題綱如下：

1. Basic technologies: protocols, service description, service interfacing, service quality-of-service, service flows and process, service discovery, service transaction, management and policy, service security and access control, web HCI

2. Interoperability: middleware interoperability, syntactic interoperability, structural interoperability, semantic interoperability

3. Middleware and tools: service authoring, service publishing, service discovery, service brokering, service creation and assembly, service integration, service collaboration, legacy system wrapping

4. Mobile web services: web services over mobile communication platforms, location-based services, location acquisition

(二) Web Service Applications

在 web services platform 上，發展架構於 Internet 的 dynamic distributed applications，以提供用戶更方便有效的數位服務。建議題綱如下：

1. Services with rich content: auctions, airline quotes, news feeds, credit check, travel schedules, publishing
2. Transactional web services for B2B or B2C: customer relationship management, supply chain management, direct marketing, dynamic pricing
3. Business process externalization: web-based EDI, web marketing, extended supply chain, business linkages at workflow
4. Location-based web services
5. Professional service automation: opportunity management, engagement structuring, project management, resource management
6. Knowledge service automation: experience sharing, knowledge capture, knowledge accumulation, knowledge search, knowledge categorization, knowledge extraction, knowledge retrieval, intellectual property brokering, patent map, risk assessment, web decision support
7. E-government
8. Web-based learning: web-based collaborative storytelling system, virtual kids city to explore identity and values, digital learning community

(三) Semantic Web

現今的 web 快速成長，但大部分所扮演的角色是文件媒體，而非可以自動化處理的資訊媒體，因此以 XML、RDF 等技術為基礎，發展對電腦有意義的新形式 web contents 及 services，將 web 轉化成 Semantic Web 是項重要的研究議題。建議題綱如下：

Basic technologies: knowledge representation to express domain ontology, agents to collect web contents from diverse sources, process the information and exchange the results with other agents, knowledge evolution

1. Visualizing tools
2. Authoring tools
3. Converting tools
4. Consistency maintenance of ontology
5. Creation and updating of ontology
6. Collaborative ontology development
7. Value-added services by semantic web

(四) Web Intelligence

利用 AI 及 advanced information technology 於 web services ，以建構 web-based intelligent information systems。建議題綱如下：

1. Web mining
2. Web warehousing
3. Web farming
4. Web knowledge discovery
5. Web event tracking
6. Hypertext analysis and transformation
7. Web log mining

(五) Sociable Web Computing

以 web services 技術實現更豐富的社交方式和內涵，使得人際溝通及社交活動，可以打破時空的限制、文化限制、語言限制等。建議題綱如下：

1. Cyber Clubs：以 web services 方式，動態組構俱樂部提供活動服務和活動場所。俱樂部成員可以在俱樂部的虛擬活動空間中自由游走、交談社交或參加團體活動。

2. Virtual Fashion and Cultural Dispersion：以 web service platform 建構成一個新興的社群平台，使其文化傳播力或流行感染力，皆高於傳統社群。

3. Universal Chat Circles：提供 universal language translation、synchronous text and audio conversation、dynamic grouping of conversational clusters、visual archival chat logs 等機能，以創造一個可以跨越國界、跨越語言的 online discussions 理想環境。

4. Live Web：讓 web users 可看到其它 users 的 real-time activities，進而了解 web population 的 behavior patterns，以增強 web community 的互動性和臨場感。

(六) Other Emerging Web Service Technologies

除了以上規劃重點和題綱外，其他前瞻性的 Web Service Technologies，也歡迎研提。

三、申請注意事項

計畫執行日期從 91 年 8 月 1 日開始。申請截止日期與 91 年度專題研究計畫相同，承辦單位為工程處資訊學門。計畫書請註明係「Web Technologies 研究計畫」。

本計畫每年將舉辦兩場計畫交流會或研討會，計畫主持人有親自參加並發表之義務。請申請人在計畫書中附同意親自參加的聲明書。

四、聯絡方式

若對本計畫有所疑問，請洽工程處王仁傑先生(tel: 27377525) (email:jcwang@nsc.gov.tw) 或台大資工系項潔教授(tel: 23625336 ext.323) (email:hsiang@csie.ntu.edu.tw)

國科會 Semantic Web 研討會

時間：中華民國 91 年 9 月 20 日（星期五）09：00~17：00

地點：中央研究院資訊科學研究所一樓演講廳

主辦單位：國科會 Web Technologies 規劃推動小組、

中央研究院資訊科學研究所

協辦單位：人工智慧學會、計算語言學學會、資訊學會、

數位典藏技術推動委員會

國科會 Semantic Web 研討會

Semantic Web 是新一代的網路資訊表達方式。希望在現有的 Web 上增加語意訊息，將網路資訊有效整合，提供更精確的自動化資訊服務。自 1998 年 Web 的發明人暨 W3C 主席 Tim Berners Lee 提出 Semantic Web 之後，相關研究在國際上受到相當的重視。爲了促進國內學術與產業界 Semantic Web 的發展，國科會委託中央研究院資訊科學研究所舉辦本次研討會，邀請國內相關學者專家進行技術交流，並且對 Semantic Web 的前景與挑戰進行討論。本次研討會同時由人工智慧學會、計算語言學學會、資訊學會、數位典藏技術推動委員會協辦。歡迎各界人士踴躍參加。

一、**時　　間**：中華民國 91 年 9 月 20 日（星期五）09：00 ~ 17：00

二、**地　　點**：中央研究院資訊科學研究所一樓演講廳

三、**主辦單位**：國科會 Web Technologies 規劃推動小組、中央研究院資訊科學研究所

四、**協辦單位**：人工智慧學會、計算語言學學會、資訊學會、數位典藏技術推動委員會

五、**議　　程**：

時間	內容	演講者	主持人
09:00~09:20	報到		
09:20~09:40	致詞及 Web Technologies 規劃推動說明： 項潔教授（台灣大學、國科會 Web Technologies 規劃推動小組召集人）		
09:40~10:40	Building Sharable Ontologies for Intelligent Agents Based on Semantic Web	蘇豐文教授（清華大學）	簡立峰博士（中研院）
10:40~11:00	中場休息		
11:00~12:00	An Introduction to Semantic Web Portal	葉慶隆教授（大同大學）	
12:00~13:30	午餐		
13:30~14:30	Semantic Web and Agents	許永真教授（台灣大學）	簡立峰博士（中研院）
14:30~14:50	中場休息		
14:50~15:50	Applications of Semantic Web	簡立峰博士、林士睿先生（中研院）	
15:50~17:00	座談會： Semantic Web 的前景與挑戰	許聞廉博士（中研院）、 郭譽申博士（中研院）、 陳銘憲教授（台灣大學）	陳克健博士（中研院）

六、**備　　註**：1. 本研討會免費參加。
2. 午餐代購餐盒費用爲新台幣 100 元整，請於當日報到時繳交。
3. 因場地限制，本研討會已於 91 年 9 月 15 日截止報名。

Building Sharable Ontology for Intelligent Agents based on Semantic Web

Von-Wun Soo
Department of Computer Science
National Tsing Hua University



Outline of the talk

- Basic concepts in Agents, ontology and Semantic Web
- Projects related to Semantic Web
  - Using Sharable Ontology to Retrieval Historical Images
  - Answer Simple Historical Questions based on Thesaurus and Ontology
- Conclusions

What is Web?

- The Web was designed as an information space,
  - useful not only for human-human communication,
  - machines would be also able to participate and help.
- Successful factors: Simple, evolution, scalability

What is Semantic Web? (According to Tim Berners-Lee)

- Knowledge Representation goes global
- Machine-understandable information
- Possible formulation of a universal Web of semantic assertions,
  - based on a common model of great generality.
- The general model is the Resource Description Framework (RDF)

What is semantic Web? (2)

- The Semantic Web is a Web that includes documents, or portions of documents, describing explicit relationships between things and containing semantic information intended for automated processing by our machines.

  According to http://swag.semanticweb.org/whatIsSW

What Semantic Web is not?

- is not Artificial Intelligence—but will provide a foundation to make the technology more feasible
- will not require every application to use expressions of arbitrary complexity
- will not require proof generation to be useful: proof validation will be enough.
- is not an exact rerun of a previous failed experiment

Why Semantic Web?

- Standardizing knowledge sharing and reusable on Web
- Interoperable (independent of devices and platforms)
- Machine readable—for possibility of intelligent processing of information

What is a software agent?

- A paradigm shift of information utilization from direct manipulation to indirect access and delegation
- A kind of middleware between information demand (client) and information supply (server)
- A software that has autonomous, personalized, adaptive, mobile, communicative, social, decision making abilities

Agents and Ontology

- Agents must have domain knowledge to solve domain-specific problems.
- Agents must have common sharable ontology to communicate and share knowledge with each other.
- The common sharable ontology must be represented in a standard format so that all software agents can understand and thus communicate with.

Agents and Semantic Web

- Semantic Web provides the structure for meaningful content of Web pages, so that software agents roaming from page to page will carry out sophisticated tasks.
  - An agent coming to a clinic's web page will know Dr. Henry works at the clinic on Monday, Wednesday and Friday without having the full intelligence to understand the text...
  - of course the assumption is Dr. Henry make the page using a off-the-shelf tool, as well as the resources listed on the Physical Therapy Association's site.

Knowledge representation on Web

- The challenge of web is to provide a language to express both data and rules for reasoning about the data [meta-data] that allows rules from any existing knowledge representation system to be exported onto web.
- Adding logic to web means to use rules to make inference, choose actions and answer question. The logic must be powerful enough but not too complicated for agents to consider a paradox.

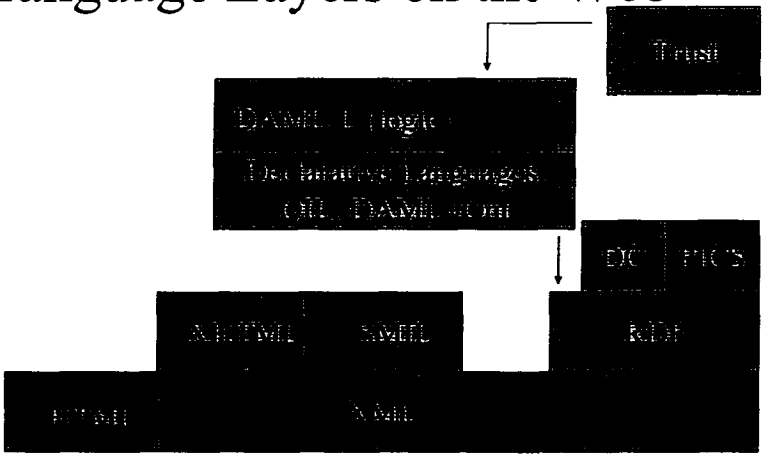
What is ontology?

- An ontology is a formal and explicit specification of shared conceptualization of a domain of interest. (T. Gruber)
  - Formal semantics
  - Consensus of terms
  - Machine readable and processible
  - Model of real world
  - Domain specific

What is Ontology?(2)

- Generalization of
  - Entity relationship diagrams
  - Object database schemas
  - Taxonomies
  - Thesauri
- Conceptualization contains phenomena like
  - Concepts/classes/frames/entity types
  - Constraints
  - Axioms, rules

Language Layers on the Web



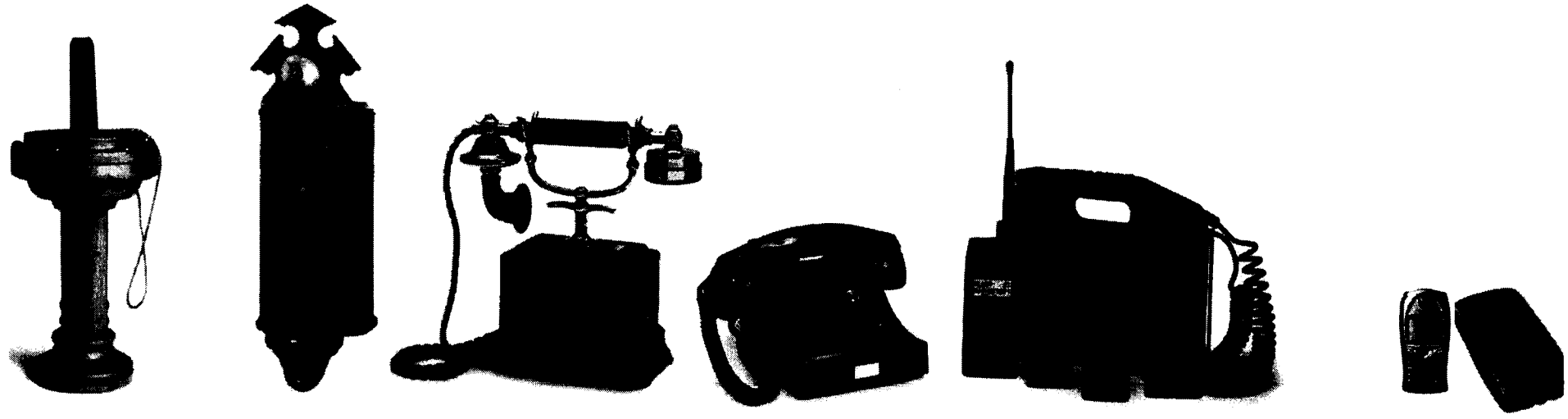
Semantic web infrastructure is built on RDF data model

Ontological languages

- Ontology modeling languages:
  - Concept Map, UML, Entity-relation Model
- Ontological languages:
  - KIF, RDF, RDF schema, DAML+OIL

Tagging documents

- Everything on semantic web is a standard hypertext tagged with "semantic" tags
- Which can be regarded as a resource



1878 1880 1910 1947 1986-96 2001-02

W@P- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

WAP Total Solution Provider.
Your Best Mobile Commence Partner

3G－我的未來不是夢

?

W@P- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.

WAP Total Solution Provider,
Your Best Mobile Commence Partner

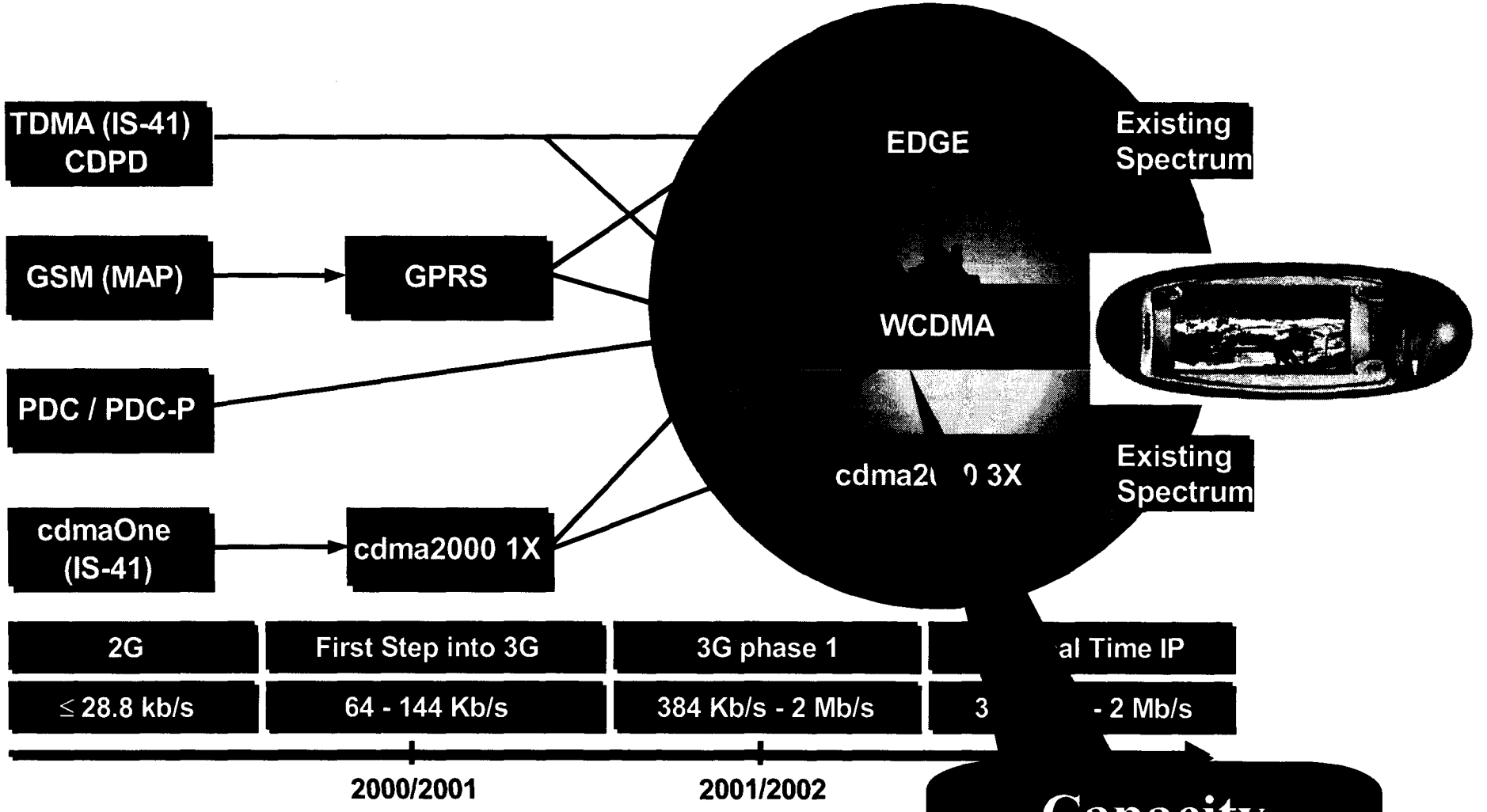
- 歐洲是第一個發放 3 G 執照的地區，但截至目前為止，所有業者對於 3G 通訊基礎建設，幾乎呈現按兵不動狀況。探究其主因為通訊業者認為過去執照拍賣價格過高，再加上現有通訊技術的影響（Wireless LAN），通訊大廠紛紛延緩 3 G 通訊基礎建設，整個 3 G 技術的成熟時間，由過去預估的 2003 年延後至 2005 年。
- 如法國雖採取評審制，但是由於執照費用訂定過高，且規定於前兩年需繳交大部分執照費用，使得四張執照只有兩家業者申請。
- 以德國為例德國平均每張執照價值高達77億美元，以15%的淨現值計算，獲得德國3G執照的業者，未來20年內，每年稅後盈餘必須達到10億美元以上，平均每人負擔3G執照成本金額高達新台幣19,100元。
- 以英國電信為例，由於在3G執照上的過度投資，2001年第一季虧損28億英鎊，為民營化17年後的首度虧損

Copy right 2002 e-Trade Info.Do not copy without Permission

- 我國第三代行動通訊業務（3G）電子競價作業，由六家公司爭取五張執照，經過19天180回合的激烈競標，在世界全通電信公司累計棄權四次，依據第三代行動通信業務管理規則第二十八條第一項規定，廢止其競標資格後，整體競價程序終告結束，3G執照確定由遠致、聯邦、台灣大哥大、中華電信和亞太行動寬頻五家公司獲得，總標金為488.99億元，較交通部公告競價底價總金額336億元，多出152.99億元。(2002/2/8)

- 在交通部電信總局即將宣布 3G 資格標的前夕，和信電訊決策高層突然宣布退出即將舉行的 3G 執照競標，和信電訊子公司和翔寬頻電訊12月27日正式向交通部電信總局提出退出 3 G執照競標的文件。 鑑於目前景氣低迷及相對於台灣行動通訊市場過多的3G執照發照數目，和信電訊經營團隊及其所有董事皆認為未來3G 市場之過度激烈競爭是可預期的。退出3G執照競標，可使和信電訊更專注於現有GPRS、igogo服務的推廣以及明年第二季在其外資夥伴NTT DoCoMo協助下推出的i-mode服務，全力推動台灣行動上網市場的發展。

Copy right 2002 e-Trade Info.Do not copy without Permission



W@P- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

Info.
n Provider.
Your Best Mobile Commence Partner

❑GPRS(General Packet Radio Service)服務特色

- 以封包傳輸方式，高傳輸速度(現階段20~50kbps)
- 不需撥接，只要一開機，隨時都處於連網狀態中
- GGSN(APN-Access Point Name)提供對Internet開放式連結
- 以資料傳輸量計費，較為合理

❑隨時上網 (Always - Connect)

❑傳輸更快速(理論值超過100Kbps)

❑交易更安全

- 可自行建置GPRS ready 的WAP Gateway並具Security功能

❑服務更便宜

- CHT: 0.025元/128 Bytes; KGT: 0.02元/128 Bytes

Copy right 2002 e-Trade Info.Do not copy without Permission

行動通訊服務的趨勢

世界變了 你呢?

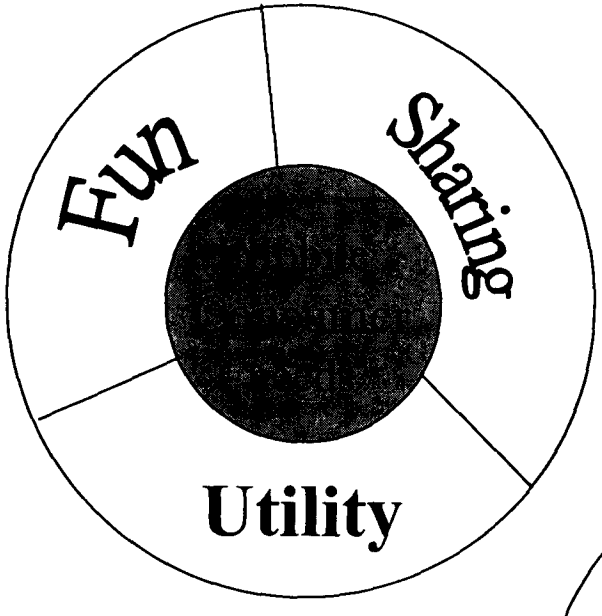
W@P- GPRS- 3G

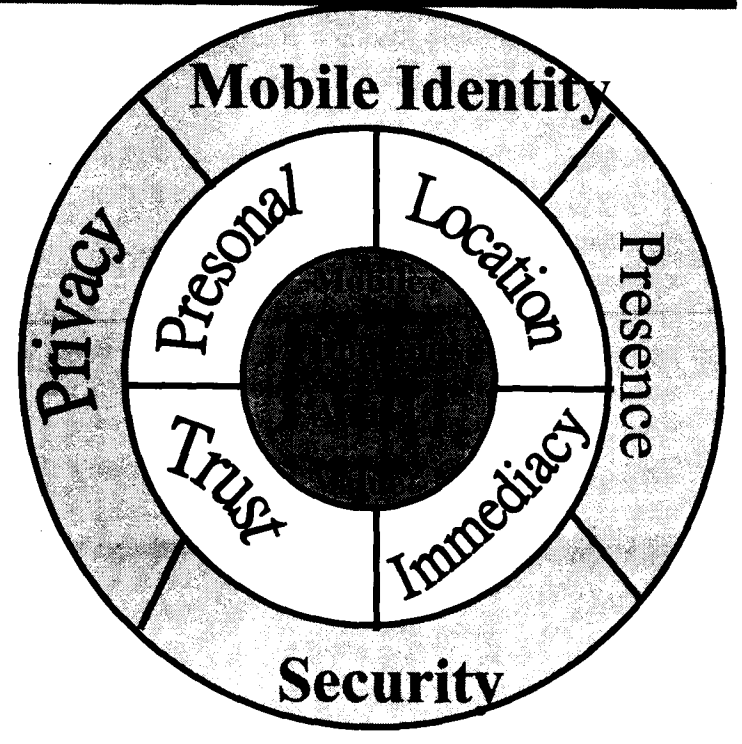
Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.

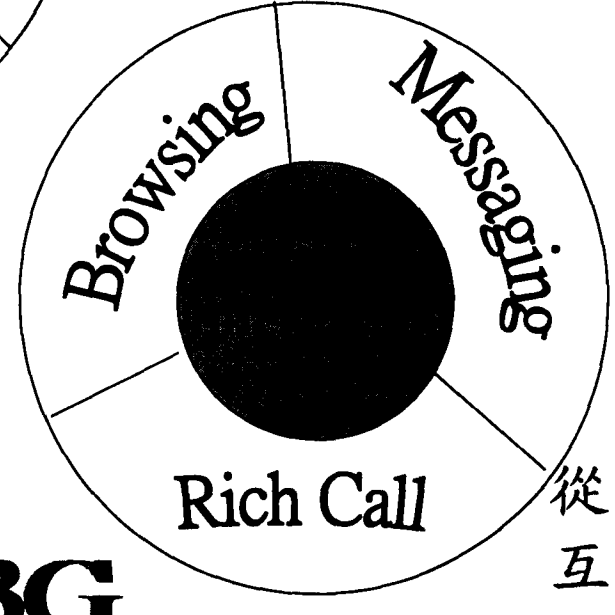
WAP Total Solution Provider,
Your Best Mobile Commence Partner

基本需求





從個人及行動中
產生附加價值



從豐富及一致性的
互動產生附加價值

W@P- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.
WAP Total Solution Provider.
Your Best Mobile Commence Partner

體育

通訊錄

自動語音系統

新聞

媒體焦點

信息整合

電子郵件

語音信箱

時辰表

小組協調

帳務

金融業務

銀行帳戶查詢



交通資訊

股市交易

定位系統

汽車派遣系統

傳真

賭博

遊戲

票務系統

月曆

告知功能

檔案傳輸

國際網路

氣象報告

WAP- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

fo.

WAP Total Solution Provider.
Your Best Mobile Commence Partner

❑Mobile Finance

- Bill Payment
- Financial Transactions
- Trading
- Banking Services

❑Mobile Intranet

- Email
- Access to corporate database
- Sales force automation

❑Mobile Commerce

- Shopping
- Ticketing
- Betting
- Information Services

❑Mobile Customer Care

- Profile change
- Service activation/deactivation

❑Mobile Entertainment

- Gaming
- Betting
- Dating

Personalization

Time-Base

Location Base

WAP - GPRS - 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.
WAP Total Solution Provider,
Your Best Mobile Commence Partner

Identifiers: Uniform Resource Identifier (URI)

- All subjects and objects in web are represented by a URI just as a link in a page
- An URL is a most common type of URI

Documents: Extensible Markup Language (XML)

- I just got a new pet dog. [An English Sentence]
- In XML:

```
<sentence><person
href="http://aaronsw.com/">I</person>
just got a new pet
<animal>dog</animal>.</sentence>
```

- Tags
- A full set of tags (opening and closing) and their content is called an element
- Descriptions such as href="http://aaaronsw.com/ are called attributes

- 輕巧化(**Light weight & long battery life**)
- 雙頻(模)/多頻(模)/**Global roaming/Inter/IntraNetworking**
  - **DECT/GSM 900/1800/1900**
  - **AMPS/D-AMPS/CDMA**
  - **GSM/CDMA/Satellite(LEO)**
- Internet Ready; High-Speed Data; IP-Base; Security
- 語音; 文字語言本土化 **(Localization)**
- Advance GUI
  - **Advanced handwriting input & recogni**
  - **Advanced VR and voice-activated diali**
  - **Image capability(e.g., Video phone)**
- 整合服務 (Integrated Services)
  - 娛樂功能 **(Entertainment)**
  - 資訊服務 **(Information/Messaging service)**
  - 多媒體 **(Multimedia)**
  - 個人資訊管理 **(Personal Information Management)**
  - 電子商務 **(E-Commerce)**
  - 個人教育**(Personal Education)**



W@P GPRS 3G

Copy right 2

網路銀行都沒起來，行動銀行有用嗎?

Copy right 2002 e-Trade Info.Do not copy without Permission

- 根據iamasia針對網友進行線上肖費金融的使用調查顯示，台灣區已有24%的網友使用網路銀行理財服務。在網路銀行知名度上，以中國信託、台新銀行與花旗銀行為前三名，但仍有三成的網友不知道哪些銀行提供網路銀行服務。而曾經使用過網路銀行的網友（120萬）中，中國信託的市佔率（41%）居冠，其次為富邦銀行與台新銀行。
- 感到滿意的部分在於網路銀行：節省往來銀行的時間（39%）、使用方便（19%）與完善的服務項目（10%）。可見節省時間與使用方便構成網路銀行最大的優勢。
- 而網友最常使用網路銀行的服務為：餘額查詢、交易明細與即時轉帳等。
- 非網路銀行使用者則表示，安全性考量與不習慣使用網路銀行是進入網路銀行的主要障礙。

Copy right 2002 e-Trade Info.Do not copy without Permission

❑Internet Banking

- **Destination not Portal**

❑滿意的部分

- 節省往來銀行的時間（**39%**）、使用方便（**19%**）與完善的服務項目（**10%**）

❑最常使用的服務為：

- 餘額查詢、交易明細與即時轉帳等

❑非網路銀行使用者

- 安全性考量與不習慣使用網路銀行

❑Mobile Banking (was)

- **Portal (Carrier) not Destination**
- **STK 僅限於特定Banks & Carriers**

❑滿意的部分

- 節省往來銀行的時間(但是花通訊費貴)、似乎沒有比較方便

❑最常使用的服務為：

- 短訊通知
- 餘額查詢、即時轉帳等 **(??)**

❑非行動銀行使用者

- 不會用**WAP**手機、連線費時、費用高、安全性考量與不習慣使用

Copy right 2002 [redacted] Without Permission

Mobile Banking
過去、現在、未來

Copy right 2002 e-Trade Info.Do not copy without Permission

❑單一大哥大、單一銀行STK (SIM Tool Kit)

- 優點
  - 預先設定選項清單方便操作
  - 專用封閉系統傳輸安全性高
- 缺點
  - 一支手機，僅能服務一家銀行
  - 需要換卡、服務不能修改

❑GSM/WAP (撥接式)

– 利用大哥大WAP Gateway

- 優點
  - 連線後可取得較多服務(上網瀏覽)
- 缺點
  - 費用高、連線過久、 人機介面複雜操作不易、有PIN明碼曝光問題

Copy right 2002 e-Trade Info.Do not copy without Permission

❑ 共用式STK (財金系統)

- 優點
  - 安全、多家銀行、多家大哥大(現在中華電信及泛亞)
- 缺點
  - 需要換卡、服務不能增加修改
  - 僅能服務中華電信客戶(中華電信emome卡系統)
  - 需要換卡、服務不能修改

❑ GPRS/WAP (封包傳輸式) –採大哥大WAP Gateway

- 優點
  - GPRS連線(Always on)後可取得較多服務、以封包計費(便宜)
- 缺點
  - 有PIN明碼曝光問題(大哥大端)、GPRS手機及使用尚未普及

❑ 銀行自建或共用GPRS/WAP Gateway

- 優點
  - 安全、 GPRS連線(Always on)後可取得較多服務、以封包計費(便宜)
- 缺點
  - GPRS手機及使用尚未普及

Copy right 2002 e-Trade Info.Do not copy without Permission

❑帳務類交易

- 即時轉帳
- 預約轉帳
- 預約轉帳取消

❑查詢類交易

- 餘額查詢
- 預約轉帳查詢
- 末筆轉帳查詢...

❑其他類交易

- 用戶登錄/密碼變更/金融卡掛失/信用卡掛失/
  分類通告/個別通告/帳務通告...

Copy right 2002 e-Trade Info.Do not copy without Permission

- ❑ 設定手機WAP服務過程複雜，客戶不易自行設定
  - ● 可否由銀行遠端替客戶設定?
- ❑ 撥接上網費時太長(平均30秒以上)
  - ● 能在5~10秒內嗎?
- ❑ 客戶需從各大哥大系統業者之WAP入口網站進入銀行所建置的WAP網站，多層架構操作過程複雜且費時過久
  - ● 客戶首頁能否為行動銀行首頁，或個人理財首頁?
- ❑ 完成一次交易時間過長(需3~5分鐘)，因傳輸速率 9.6Kbps
  - ● 有辦法更快嗎?
- ❑ 計時收費：完成一筆交易約需花費15~24元以上
  - ● 一筆交易 NT$1 元內，可以嗎?
- ❑ 大哥大系統業者的WAP Gateway與行動銀行的Server之間有'明碼'，交易安全有顧慮
  - ● 至少要128-bit SSL 等級，否則無法提供交易類服務!!

Copy righ

Mobile Banking

新應用模式

Copy right 2002 e-Trade Info.Do not copy without Permission

DTD (Data Type Definition)

- XML's document consists of elements with attributes
- Define element
  - <!element code (#PCDATA)>
  - <!element message (ANY)>
- Define Attribute
  - <!ATTLIST authorlist type CDATA #IMPLIED>
  - <!ATTLIST authorlist type CDATA #REQUIRED>
  - <!ATTLIST book company CDATA #FIXED "Microsoft">
  - ...

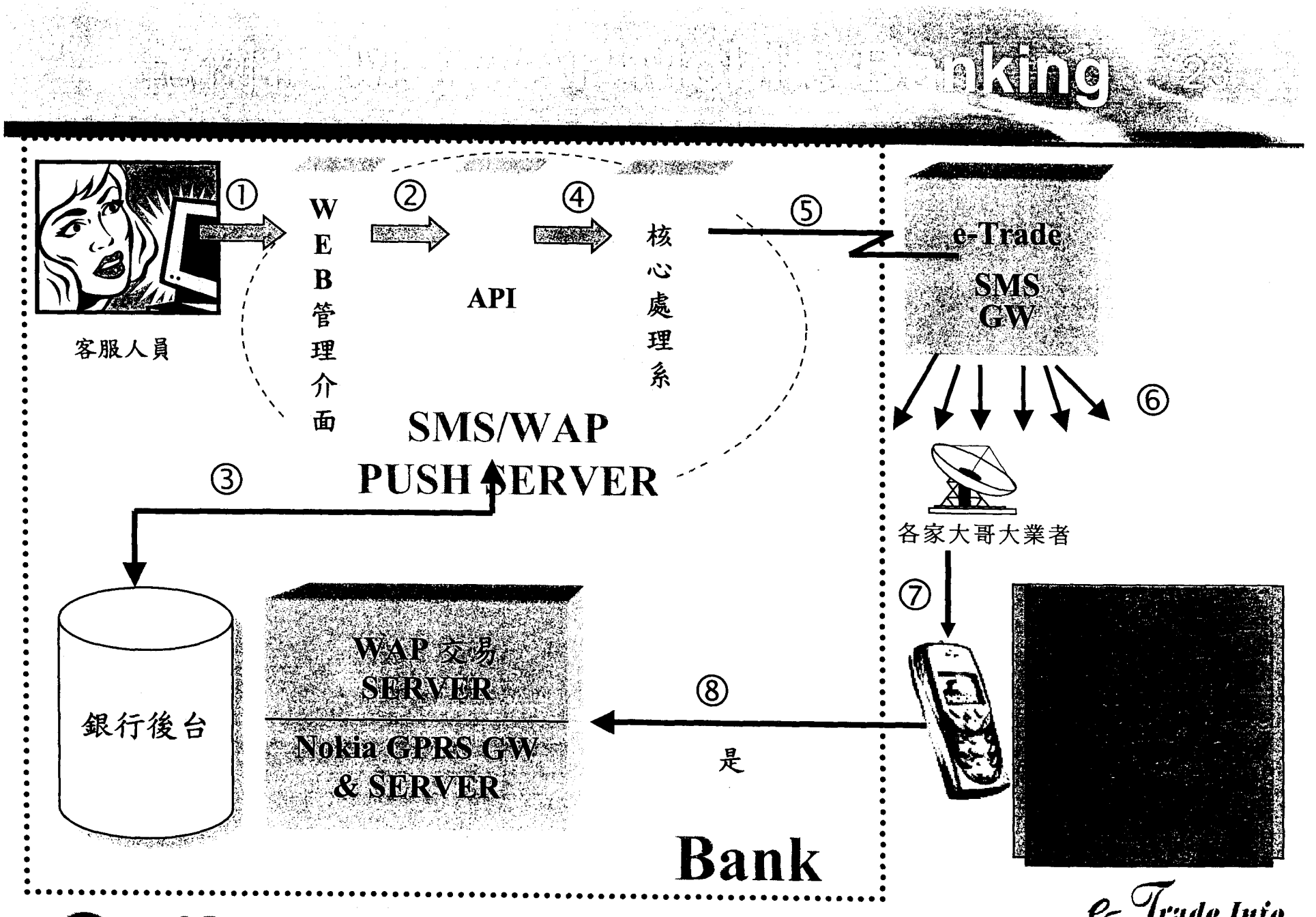
XML Schema

- A well defined XML document
  - Support more data types
  - Support name space (more extensible than XML DTD)
- Disadvantage of DTD:
  - allow user to define "ill-defined" elements

新應用模式(銀行與客戶互動)

- ❑化複雜為簡化 **(Easy to Use)**
  - **Internet Banking (上網用戶) vs. Mobile Banking(手機用戶)**
- ❑化被動為主動 **(Proactive)**
  - 由客戶主動轉換為銀行主動
- ❑化定點為隨處交易**(Anywhere)**
  - 手機成為行動**ATM**
- ❑化不知何時為定時及即時 **(Activity-Based)**
  - 定時或即時通告並傳送無線上網連結
- ❑行動銀行五大特色
  - 速度快、效率高、費用低、安全性高、個人化

Copy right 2002 e-Trade Info.Do not copy without Permission



W@P - GPRS - 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.

WAP Total Solution Provider.

Your Best Mobile Commence Partner

0 Manuals 零選單

1 Button Connect 一指上網

2 Seconds Response 兩秒回應

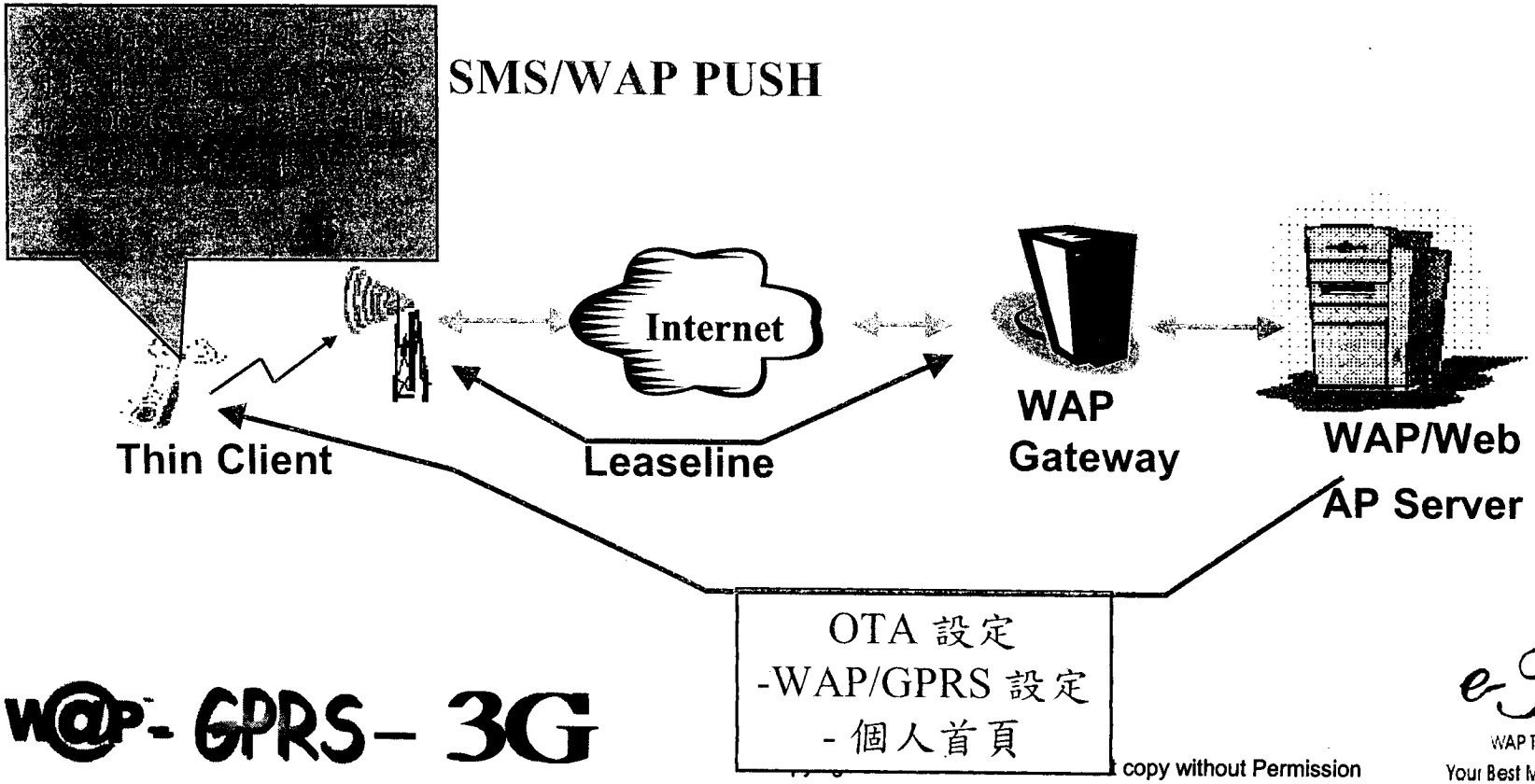
3 Clicks to Destination 三步到位

W@P - GPRS - 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.
WAP Total Solution Provider.
Your Best Mobile Commence Partner

- ❑ OTA (Over-The-Air)
- ❑ SMS/WAP PUSH
- ❑Personalized Home Page



W@P- GPRS- 3G

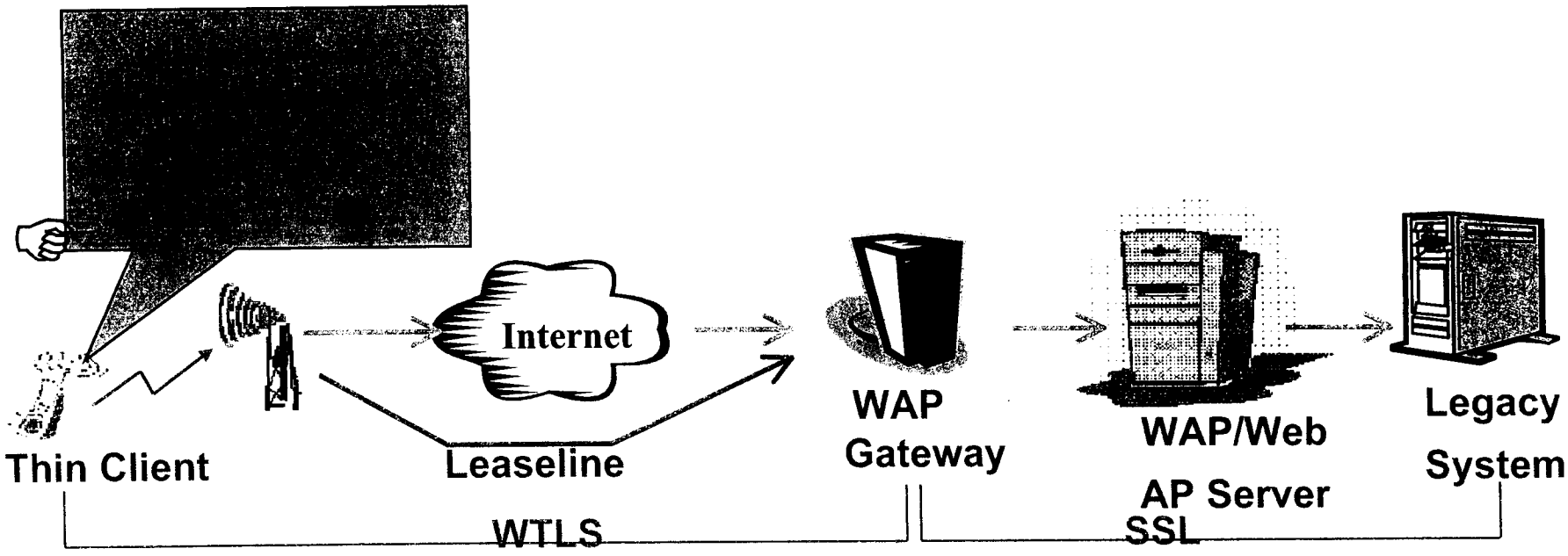
copy without Permission

e-Trade Inio.
WAP Total Solution Provider.
Your Best Mobile Commence Partner

❑ Direct Connection to Personal Page

- 個人首頁=目的網頁 (Destination)

❑WAP PUSH已提供連結(URL Hyperlink)

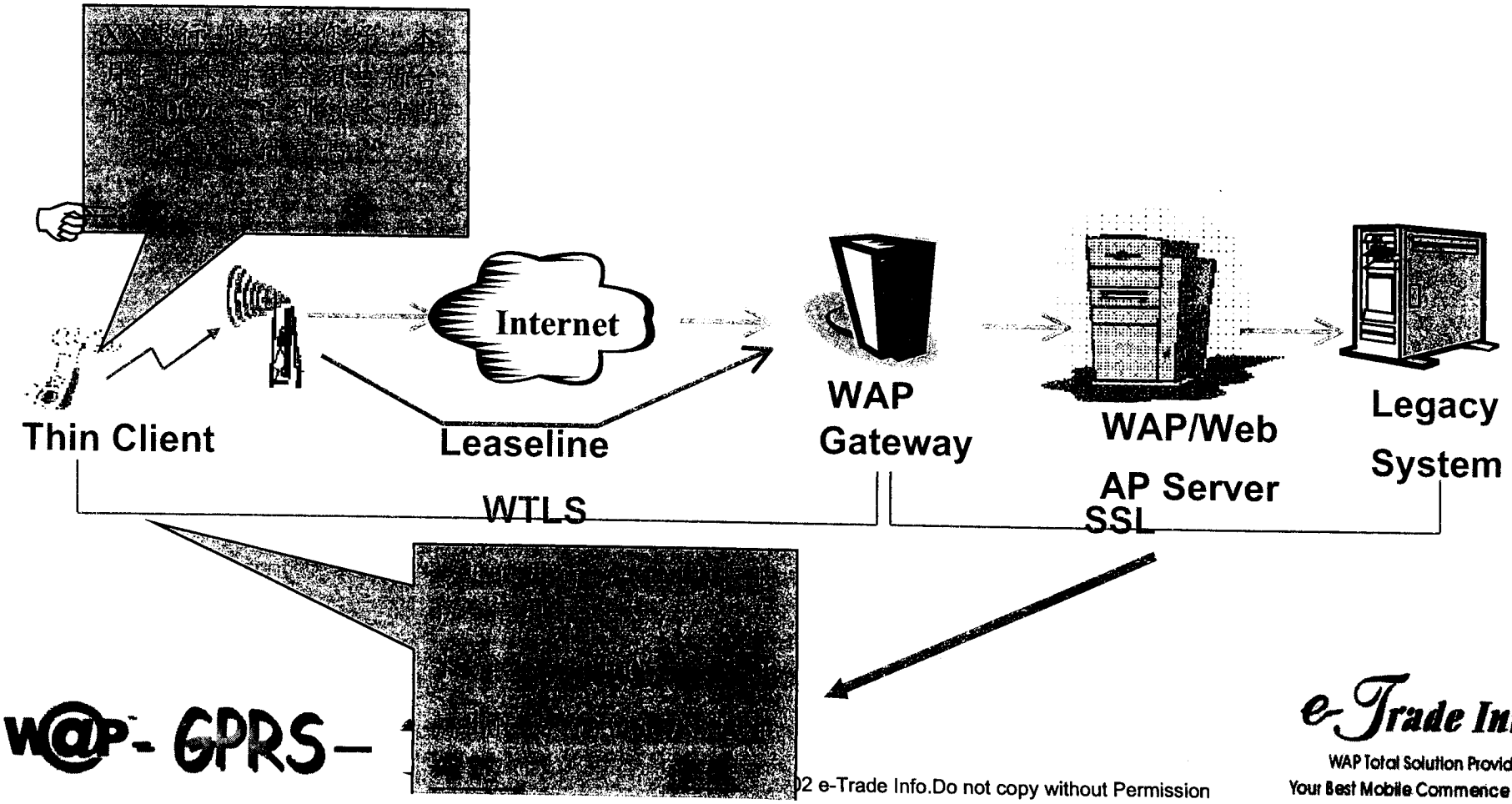


W@P- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.
WAP Total Solution Provider.
Your Best Mobile Commence Partner

- 以封包傳輸方式，高傳輸速度(現階段**20~50kbps)**
- 不需撥接，只要一上網，隨時都處於連網狀態中 **(Always On)**



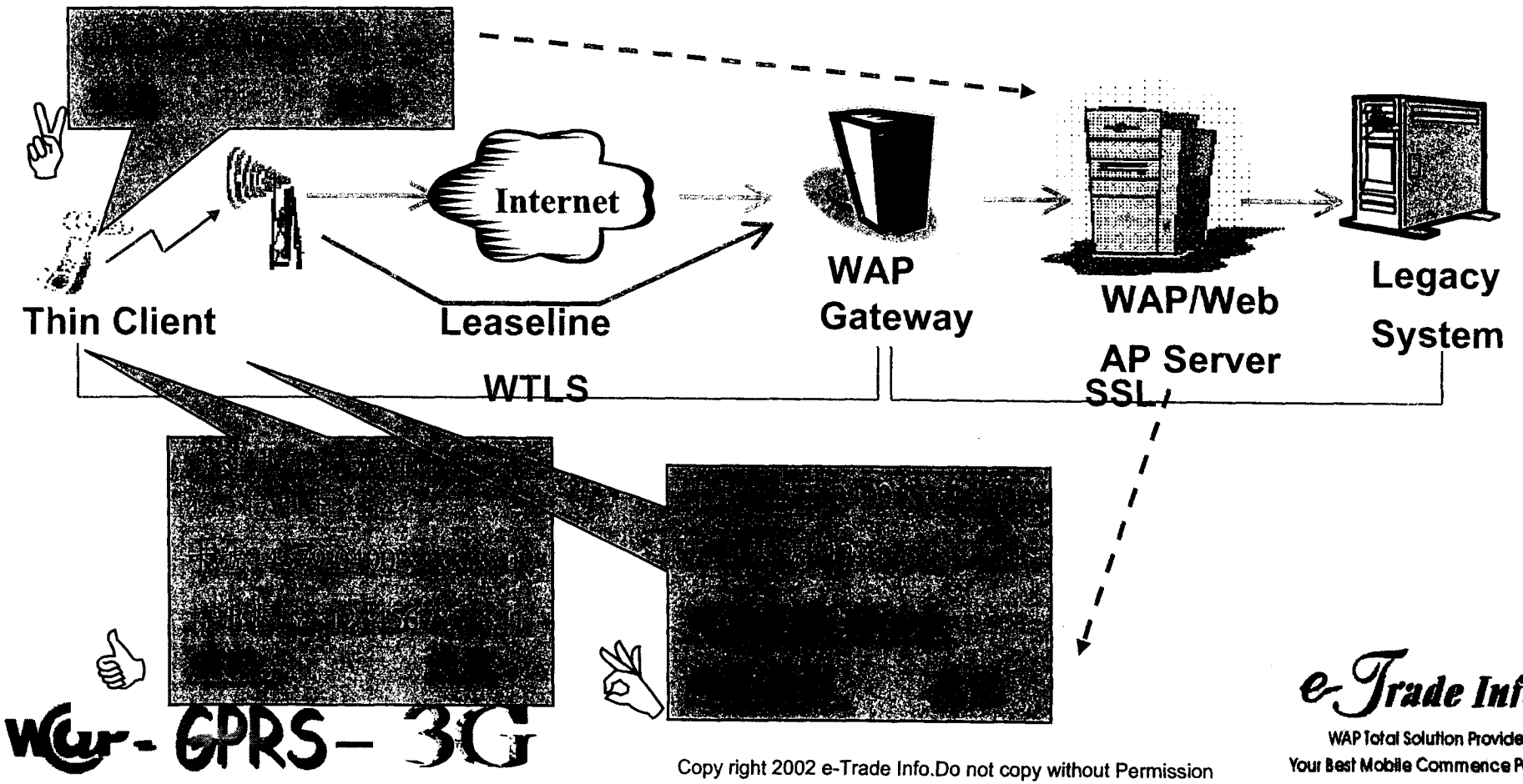
W@P - GPRS -

02 e-Trade Info.Do not copy without Permission

e-Trade Info.
WAP Total Solution Provider.
Your Best Mobile Commence Partner

❑Good service flow design

❑引導互動式的服務(客戶不覺得在上網)



WCur- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.
WAP Total Solution Provider.
Your Best Mobile Commence Partner

- ❑整合SMS/WAP Push網路平台(with info. "Push")
- ❑CA built-in on SIM = WIM (WAP1.2)
  - Wireless PKI指日可待
  - Dual-Chip or Dual-Slot
- ❑Content Aggregator & Provider to Any Devices
  - 銀行提供個人化多元、即時的理財金融資訊服務
    - 即時匯/利率報價
    - 即時財經新聞
    - 國際財經資訊
    - 各類金融商品即時報價
    - 各銀行最新活動報導
  - 整合SMS, 雙向SMS, WAP PUSH, 交易, 付款機制
  - Support PDA, WAP/GPRS phones, IVR, Fax, Web, CTI, ...
- ❑m-Commerce Platform
  - Mobile Payment services to merchants & operators

Copy right 2002 e-Trade Info.Do not copy without Permission

- 新溝通工具: 手機 = 隨身ATM+個人理財機
  - 查詢、轉帳、交易、理財、購物.....
- 新經營模式: 由客戶的主動要求服務(臨櫃、ATM 、Internet Banking 、語音銀行...)藉由個人隨身手機轉化為由銀行主動提供服務:例
  - 到期繳款通知(WAP Push)引導回轉帳
  - 累計紅利積點更換獎品通知及線上兌換
  - 個人隨身即時理財服務告知
- 新業務收入: 新種消費金融服務
  - 手機購物、支付旅遊險、.....
- 更快速服務提供: WPKI憑證的唯一及不可否認性可提供許多金融機構需"臨櫃申辦"的新業務的申請

Copy right 2002 e-Trade Info.Do not copy without Permission

未來Mobile Internet的架構

IP Convergence + Web Service

Copy right 2002 e-Trade Info.Do not copy without Permission

XML namespaces

- A namespace is a collections of names that are defined in some way.
- With XML Name Spaces(give each element and attribute a URI).
- <sentence
  xmlns=http://example.org/xml/documents/
  xmlns:c=http://animals.example.net/xmlns/>
  <c:person c:href= "http://aaronsw.com/">I</c:person>
  just got a new pet <c:animal>dog</c:animal>.
  </sentence>

XML is not the solution

- Meaning of XML-documents is intuitively clear
- But computers do not have intuition
  - Tag-names per se do not provide semantics
- DTD or XML Schema does not distinguish between objects and relations
- XML lacks a semantic model
  - Has only a "surface model", i. e. tree.



w@P- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e Info .

WAP Total Solution Provider.
Your Best Mobile Commence Partner

[redacted]

Enjoy The Day!!

Thank You Very Much

W@P- GPRS- 3G

Copy right 2002 e-Trade Info.Do not copy without Permission

e-Trade Info.

WAP Total Solution Provider,
Your Best Mobile Commence Partner

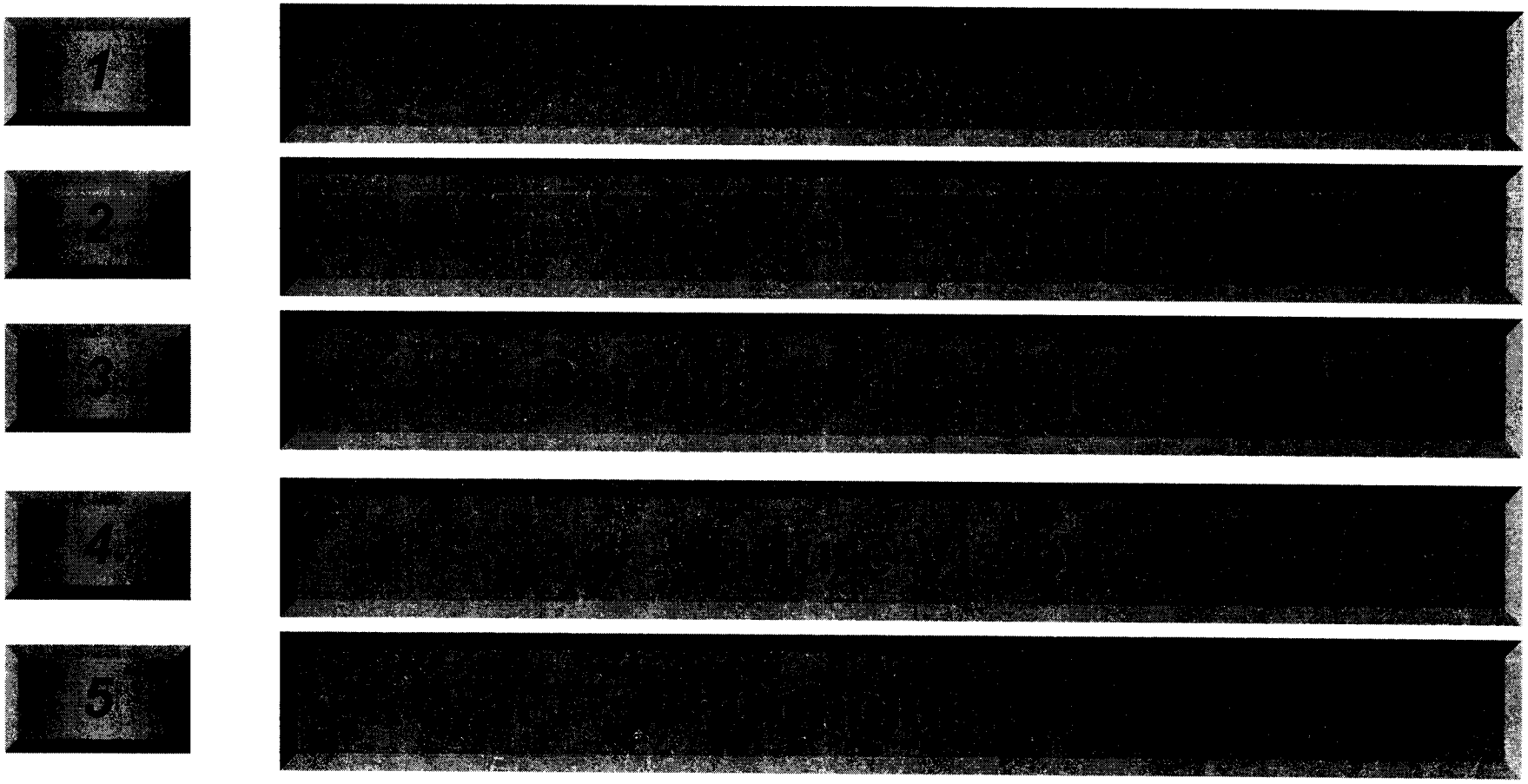
弘運科技股份有限公司
HOWIN TECHNOLOGIES CO., LTD.

無線網路服務的未來展望
Vision of Wireless Network Service

鄒伯川
副總經理
Mobile: 0958-604-818
Email: paul_chow@howin.com.tw
91年6月28日

Confidential and Proprietary

Agenda



Confidential and Proprietary

1

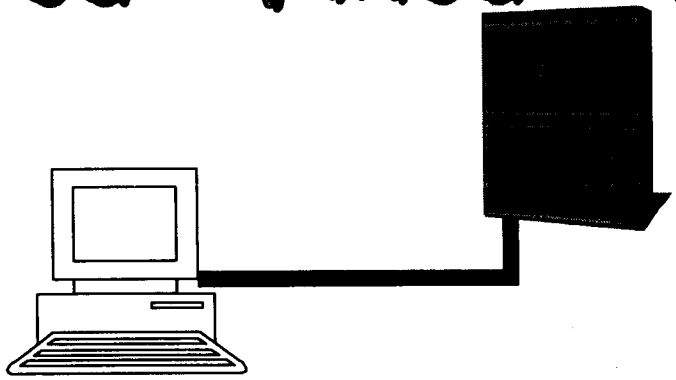
[redacted]

Confidential and Proprietary

Evolution of Communication

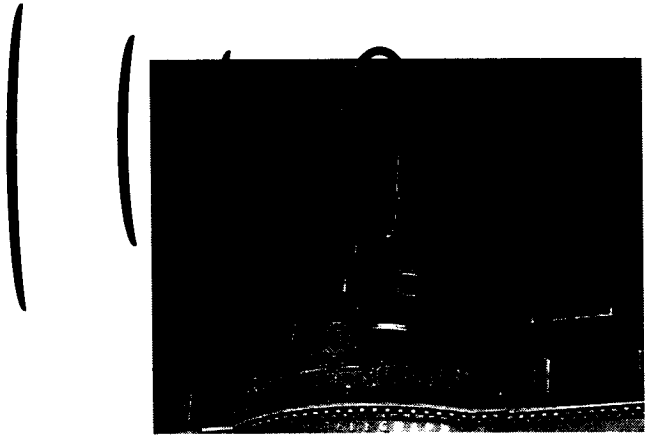
Wired Fixed Voice











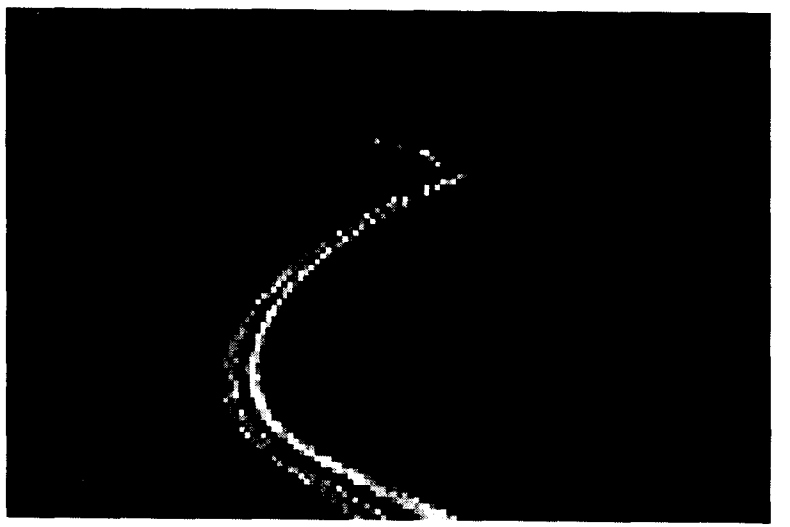


Wireless Mobile Data

Confidential and Proprietary

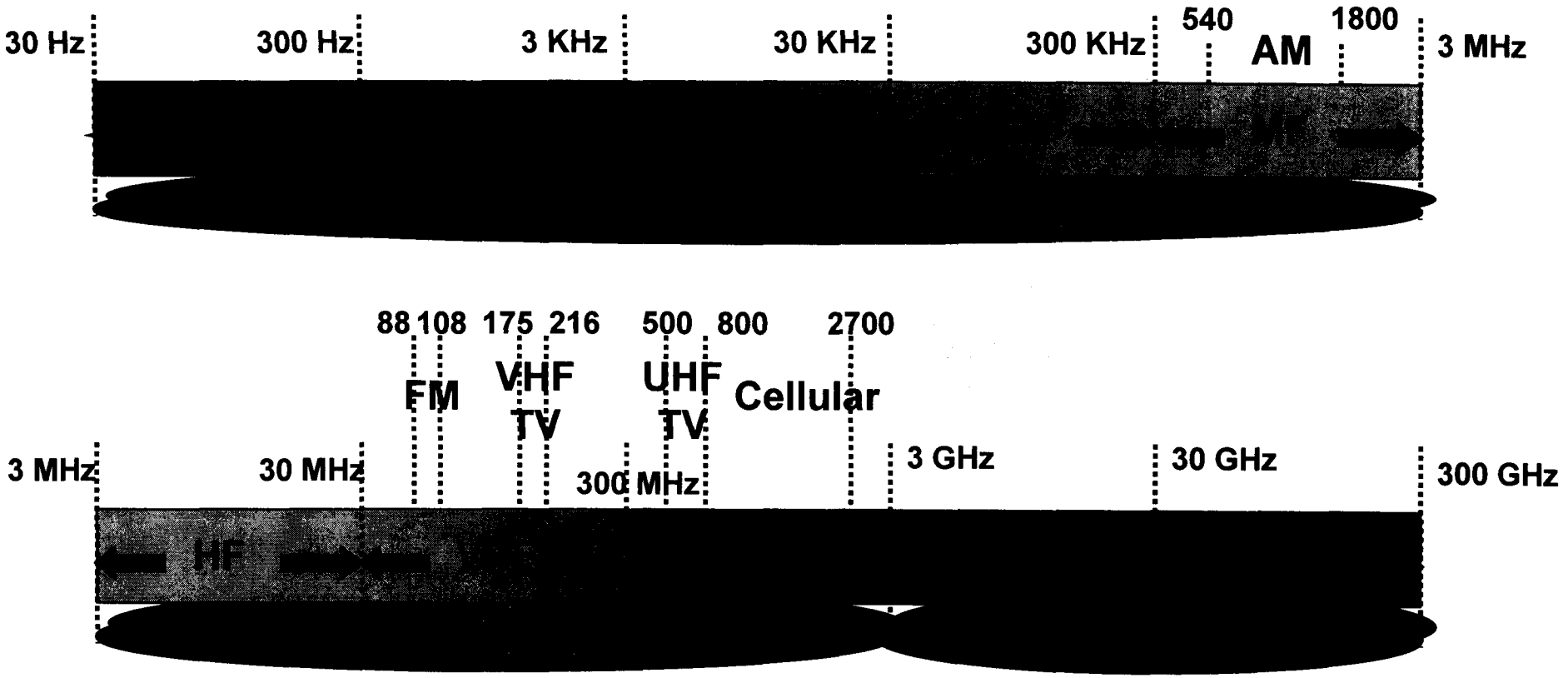
Trend of Wireless

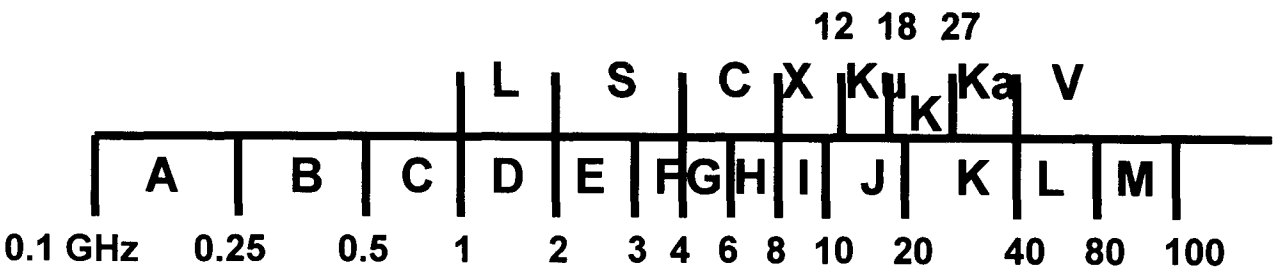
- **Why**
  - Cable-less
  - Easy to connect, access and carry
  - Provide service with less limitation
- **How**
  - Electromagnetic wave propagation
  - Antenna transmit and receive



Confidential and Proprietary

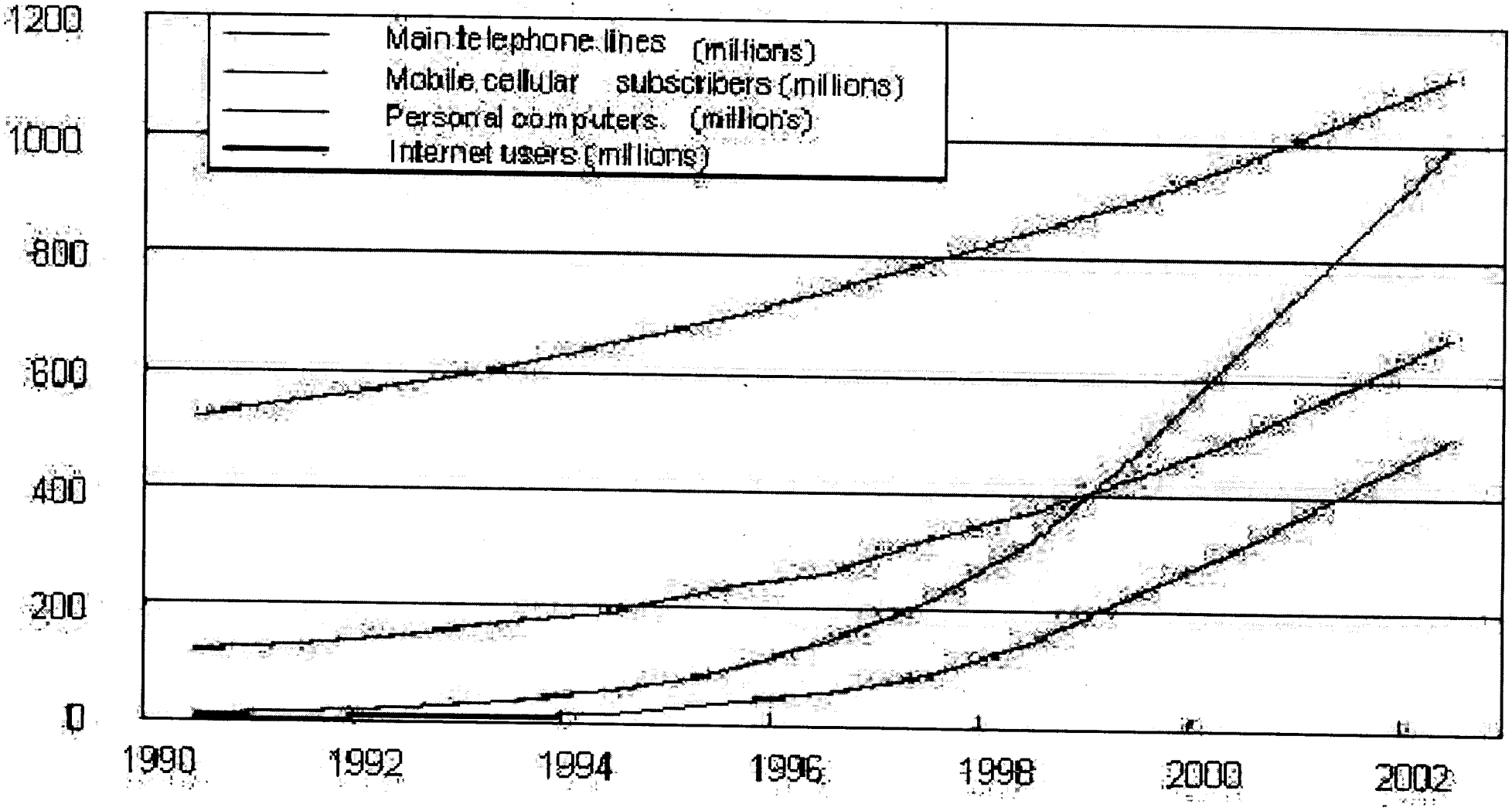
Spectrum





Confidential and Proprietary

Subscriber of Communication

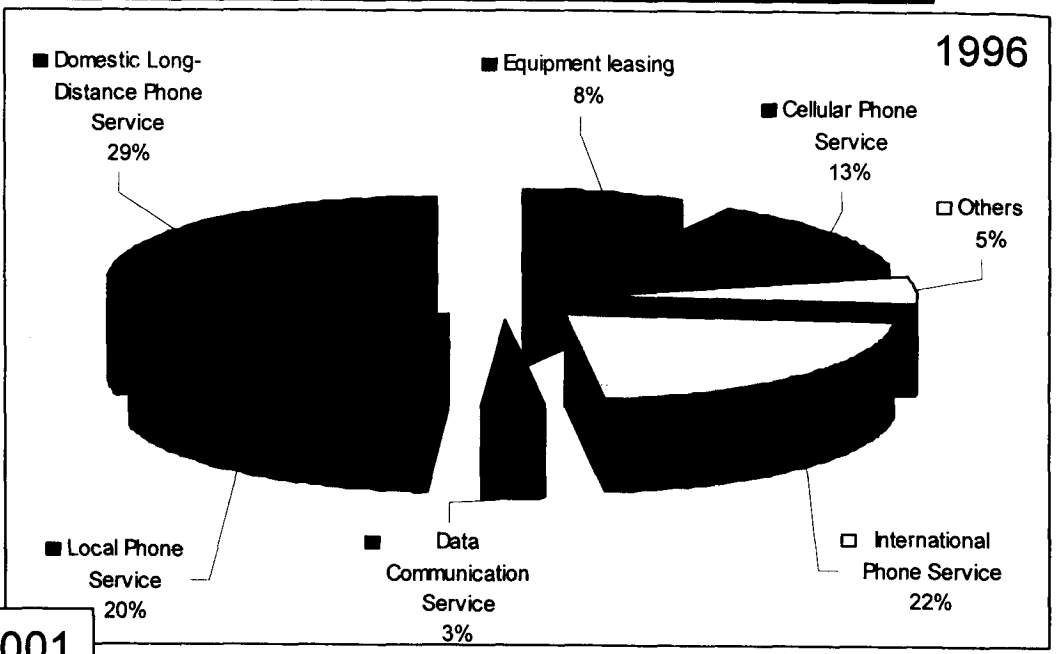


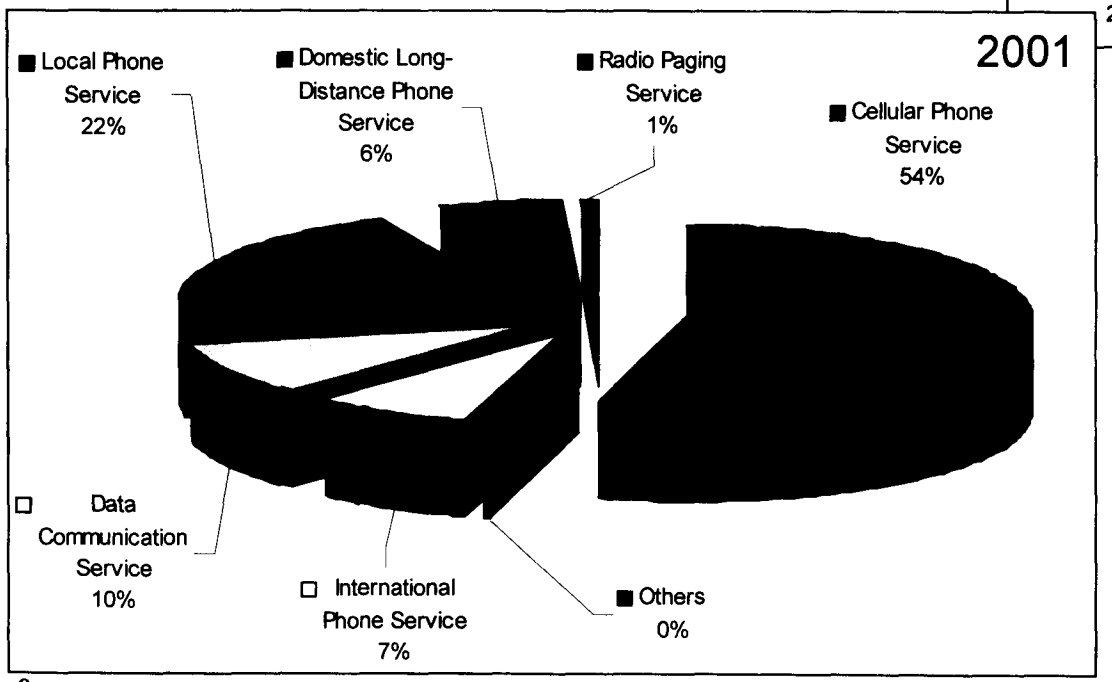
Source: ITU, 2000

Confidential and Proprietary

Revenue of Telecom Services in Taiwan

Total revenues in Taiwan's telecom market during the year 2001 approached NT$315 billion, accounting for 3.31% of gross domestic product (GDP).





Cellular phone service grow from 13% to 54.25% during 5 years.

About 66.7% of cellular phone service revenue in 2001 is coming from private telecom companies.

Source: DGT, 2002

Confidential and Proprietary

XML is not the solution(2)

```
<person>
<idn>5634</idn>
<name>W. Chen</name>
<marriedWith>
    S. Chen</marriedWith>
<gender>male</gender>
<salary>50000NT</salary>
</person>
```

```
<man idn="5634">

  <name>W. Chen</name>
  <marriedWith ref="4365"/>

<salary>1650 USD</salary>
</man>
```

Challenges: Name conflict
Value Conflict
Structure Conflicts

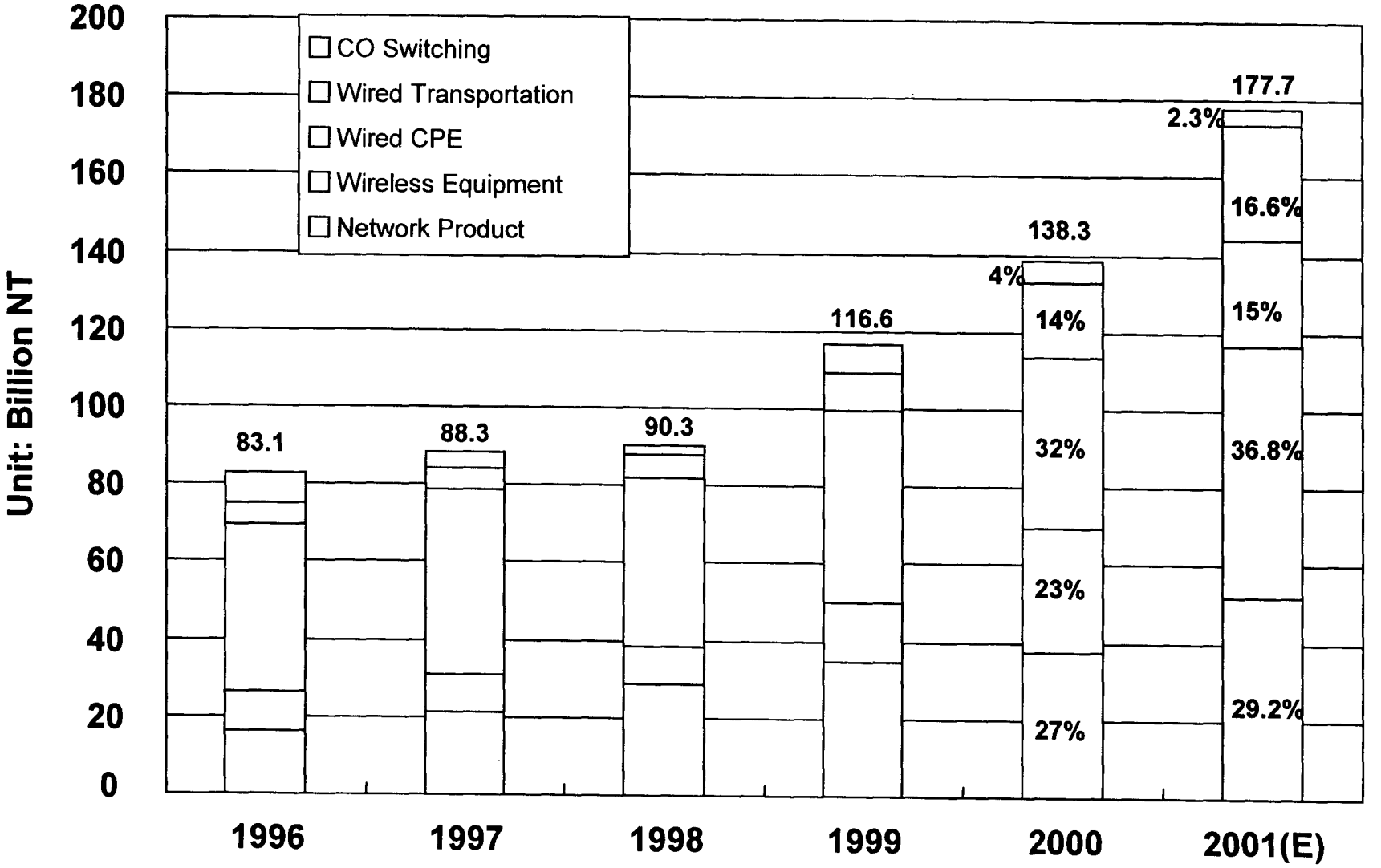
Statements: Resource Description Framework (RDF)

I really likes weaving the web.

http://aaron.com/
http://love.example.org/terms/reallylikes
http://www.w3.org/People/Berner-Lee/Weaving/

Communication Industry Trend in Taiwan

Wireless terminal is the # 1 in consumer electronic industry

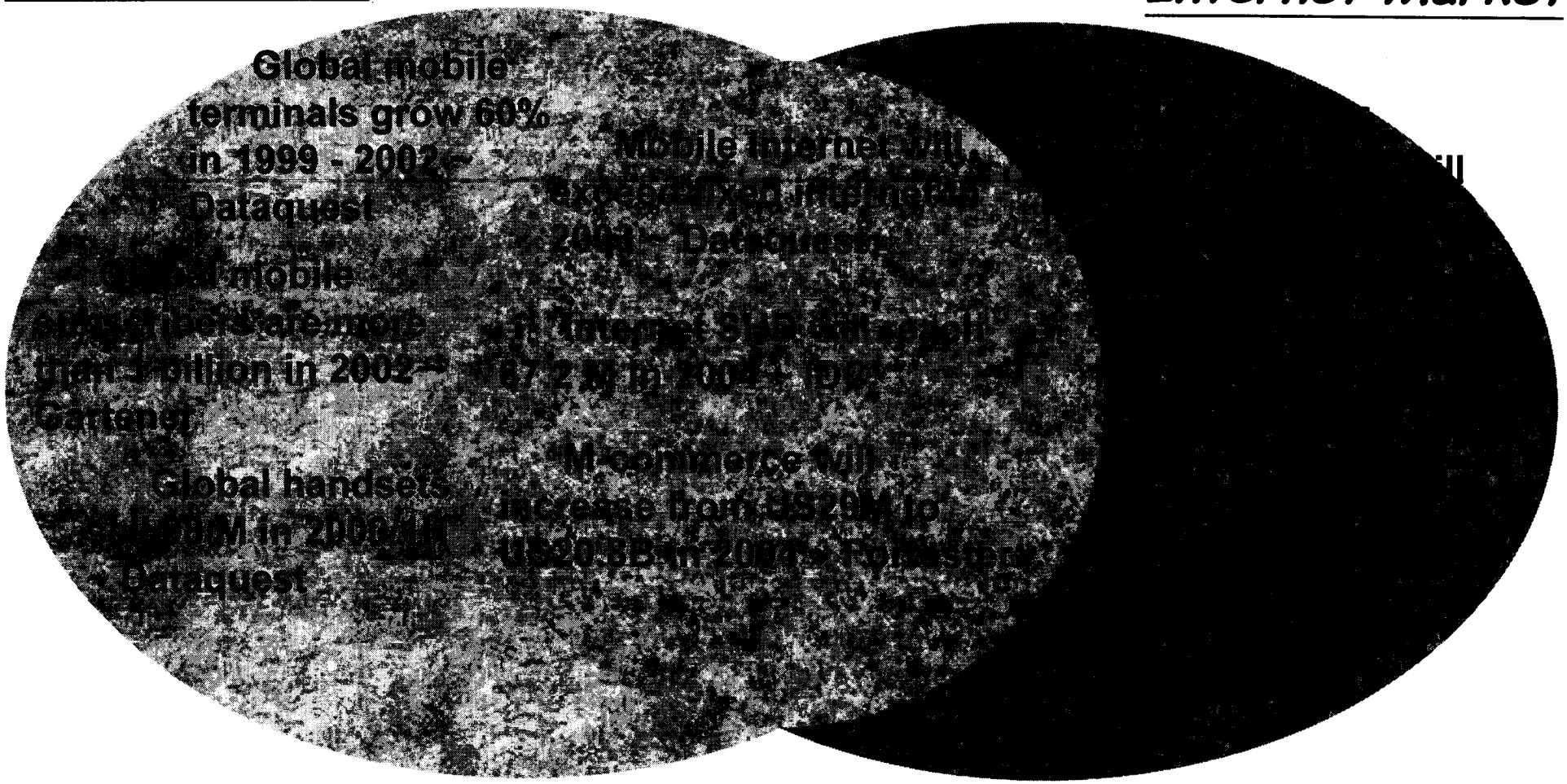


Source: Industrial Economic Knowledge Center, ITRI, 2001/11

Confidential and Proprietary

弘運科技股份有限公司
HOWIN TECHNOLOGIES CO., LTD.

Vision of Mobile Internet



Confidential and Proprietary

Market Potential

- By 2003, up to 80% of its subscribers in developed countries will use wireless data and the data will account for 20% of wireless revenue. (Source: Vodafone)

- By 2004, the number of wireless Internet users will equal the number of wired Internet users, at about 1.1 Billion. (Source: Motorola)

- By 2010, 33% of all the transactions done over the Internet could be done over cellular phones. (Source: Merrill Lynch)

Confidential and Proprietary

2

Wireless Technology

Wireless WAN: Mobile Data

Wireless LAN: 802.11

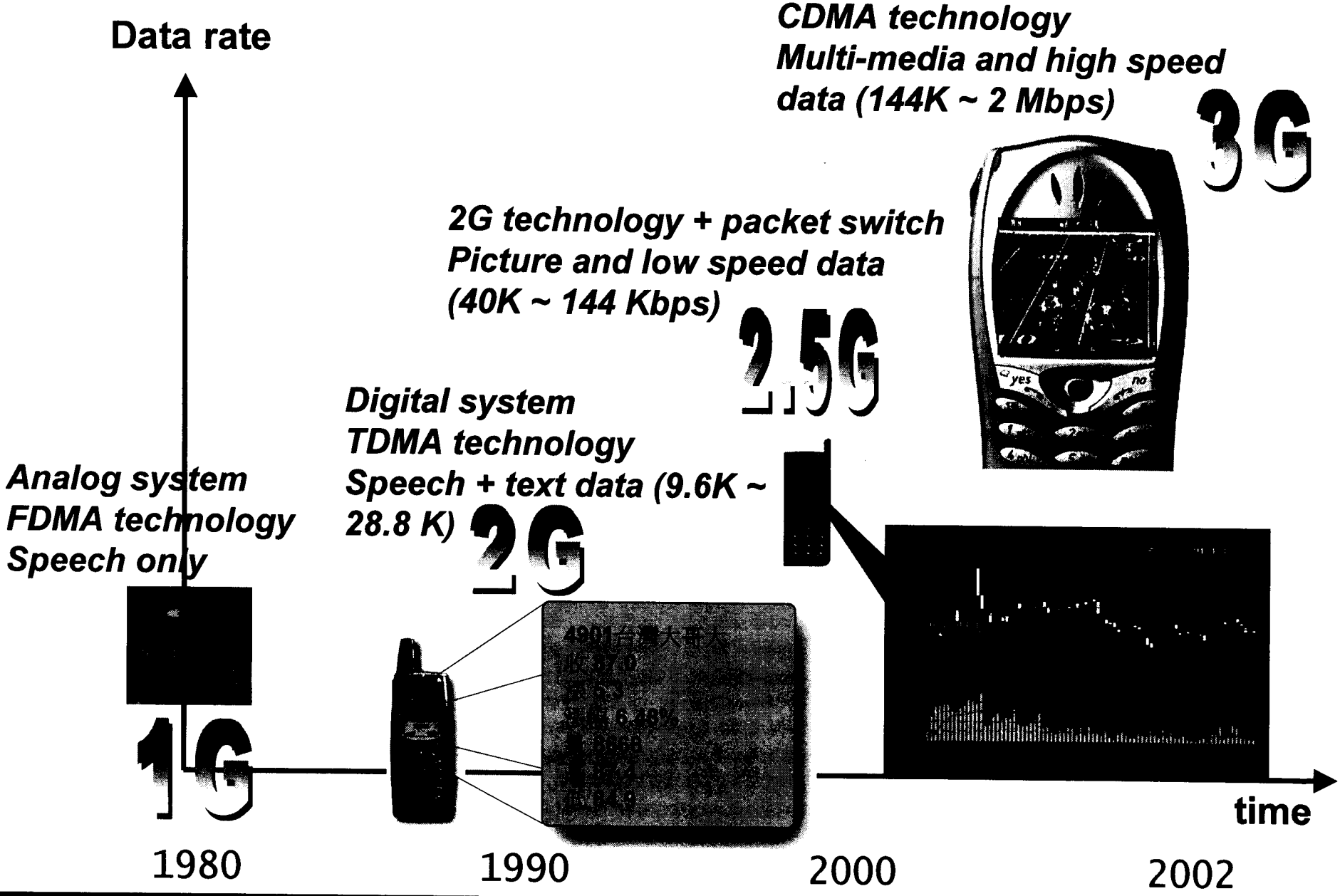
Wireless PAN: Bluetooth

Optical

Satellite

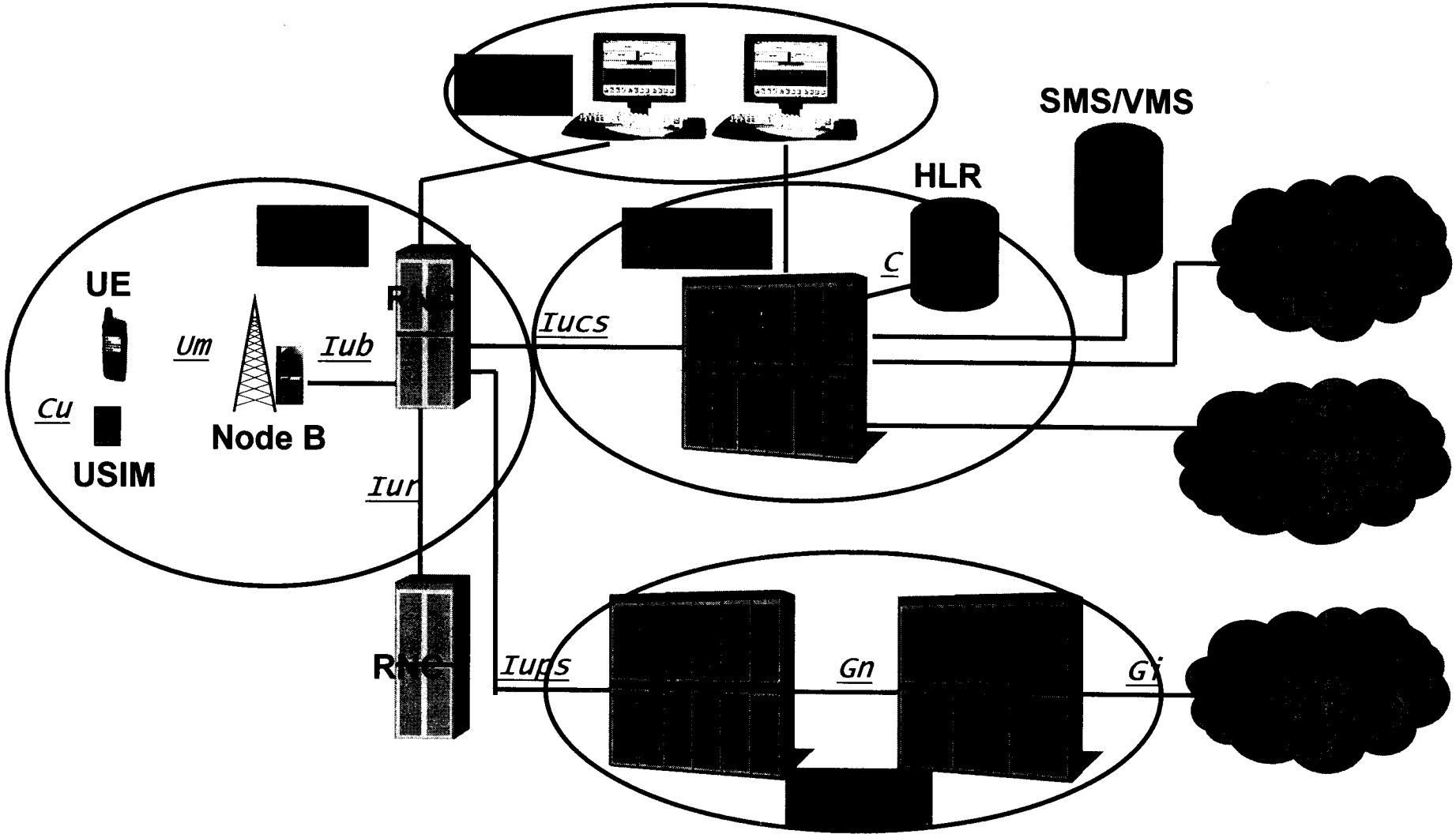
Confidential and Proprietary

3G: Improving Data Rate



Confidential and Proprietary

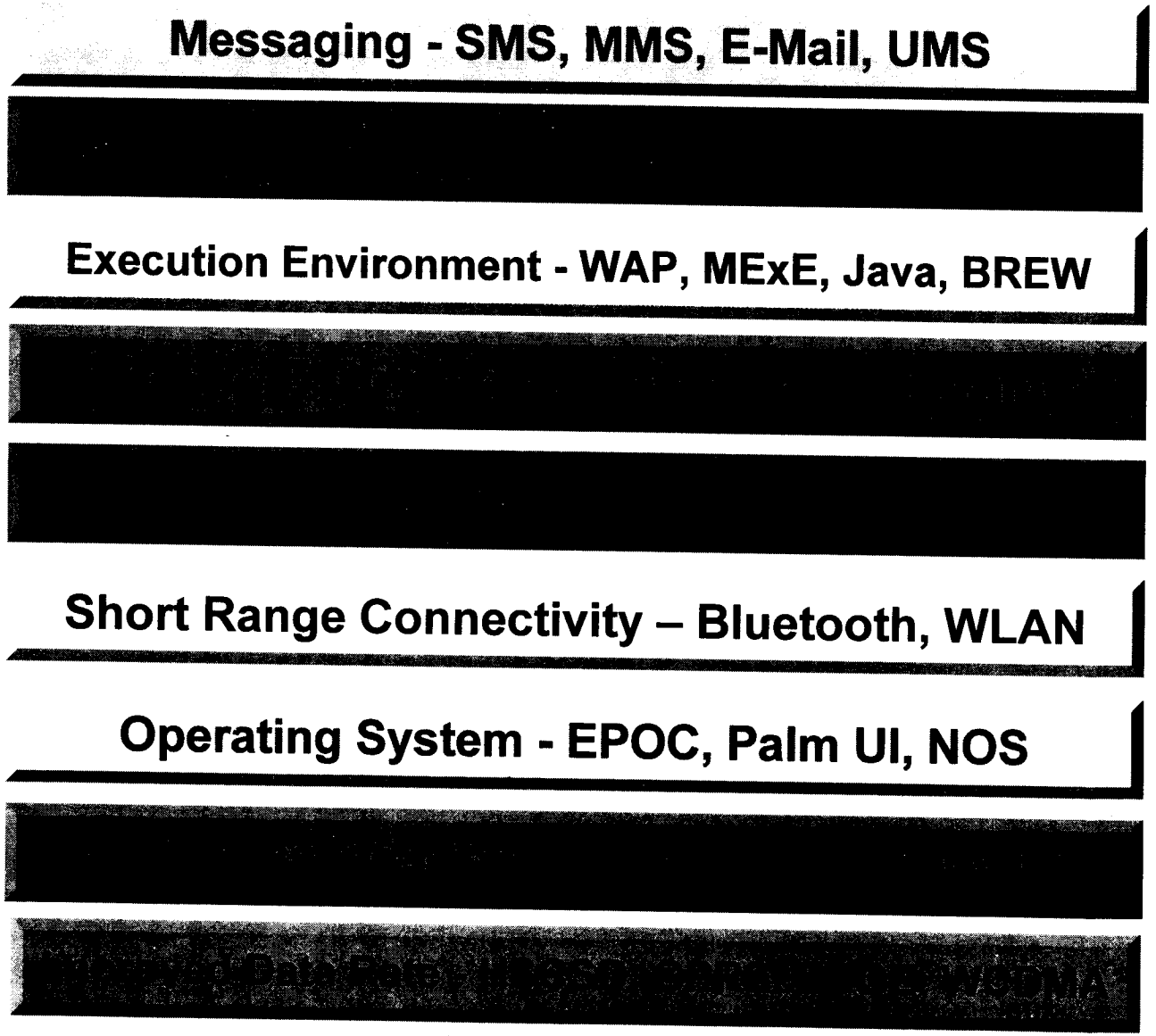
Network Topology of UMTS



Confidential and Proprietary

3G: Not Only Radio

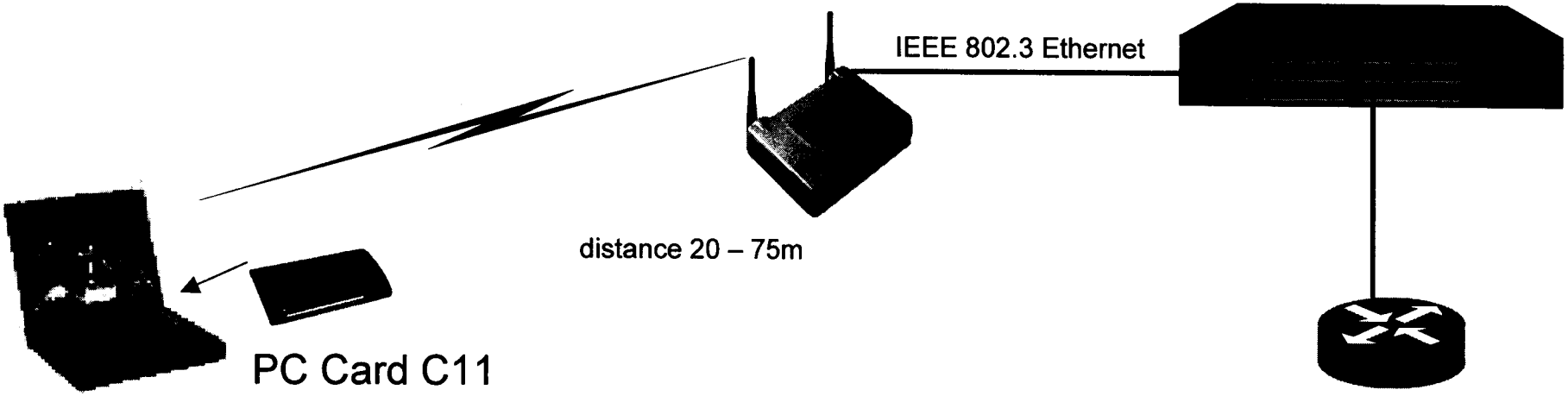
- Many technologies reach commercial maturity in the mobile environment and will exist in parallel with 3G radio
- The new digital service environment is relying on all these and more creative Wireless Data Service



Confidential and Proprietary

WLAN

- WLAN can provide high data rate access for wireless data service
- Important technologies in the mobile communication environment that build up the infrastructure for Wireless Data Service
- Uncontrolled frequency band - Quality of Service is not guaranteed
- Very suitable for local access outside operator networks - eg. company internal solutions



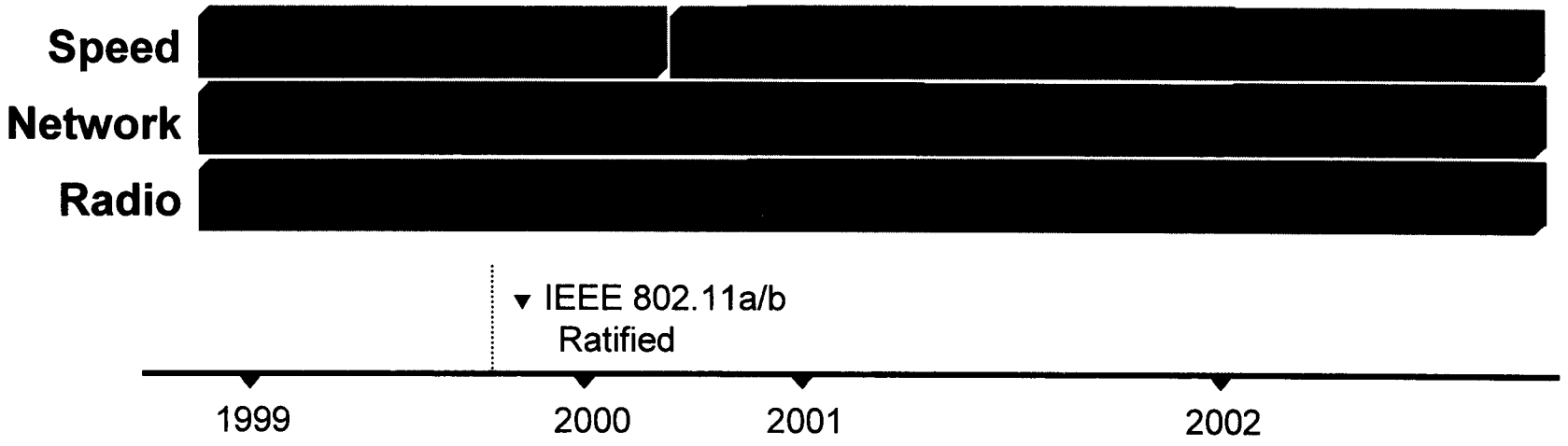
Confidential and Proprietary

	3G (WCDMA)	WLAN
Band	2 GHz	2.4 GHz, 5.8 GHz
Technology	DS-CDMA	Direct Sequence SS
Maximum Data Rate	Theoretic: 2 Mbps Practical: 384 Kbps Future: 10 Mbps	Theoretic: 11 Mbps Practical: ~ 5 Mbps Future: 54 Mbps
License Fee	Yes	No
Mobility	High	Low
Coverage	350 m ~ 1 Km	50 m ~ 100 m
Monthly Fee	0 ~ 2,200 NT (average 600 NT)	250 ~ 500 NT
Price per BS	> 3,000 KNT	< 10 KNT
Price per Terminal	~ 15,000 NT	~ 2,000 NT

Monthly fee is from Japan NTT and NTT DoCoMo

Confidential and Proprietary

Vision of WLAN



Confidential and Proprietary

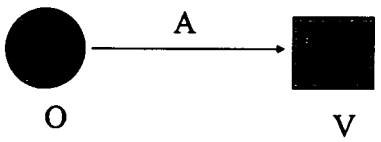
Statements: RDF(2)

```
<rdf:RDF
  xmlns:rdf=http://www.w3.org/1999/02/22-rdf-
    syntax-ns#>
  xmlns:love=http://love.example.org/terms/>
  <rdf:Description rdf:about=http://arron.com/>
    <love:reallyLikes
    rdf:recource="http://www.w3.org/People/Bern
    ers-Lee/Weaving>
    </rdf:Description>
</rdf:RDF>
```

Statements: RDF(3)

④ The basic structure of RDF is object-attribute-value

④ In terms of labeled graph: [O]-A->[V]



Problem of Wireless LAN

- Secured Corporate Access - VPN
- Authentication
  - NAI – based (user@domain + static password)
  - SIM – based (through One-Time Password over SMS)
- Administrative Integration
- WLAN access management
- Roaming (Radius – based)
- Charging (including prepaid vouchers)
- Network Management: WLAN Manager (SNMP)

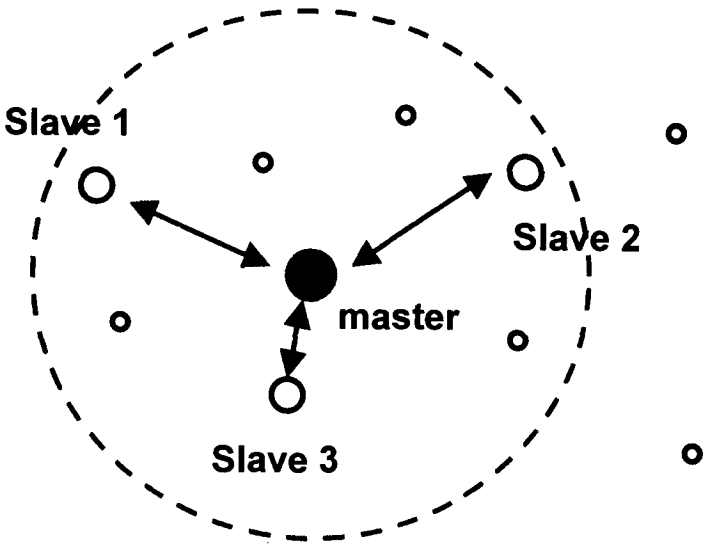
Confidential and Proprietary

WPAN

- Allows for the replacement of many proprietary cables which connect one device to another with one short-range radio link.
- Bluetooth operates in the unlicensed ISM band at 2.4 GHz is the popular solution for WPAN.
- Bluetooth can support
  - An asynchronous data channel
    - Maximally 721 Kbps while permitting 57.6 Kbps in the return direction
  - Up to three simultaneous synchronous voice channel
    - Each with 64 Kbps voice link
  - A channel with simultaneously support asynchronous data and synchronous voice



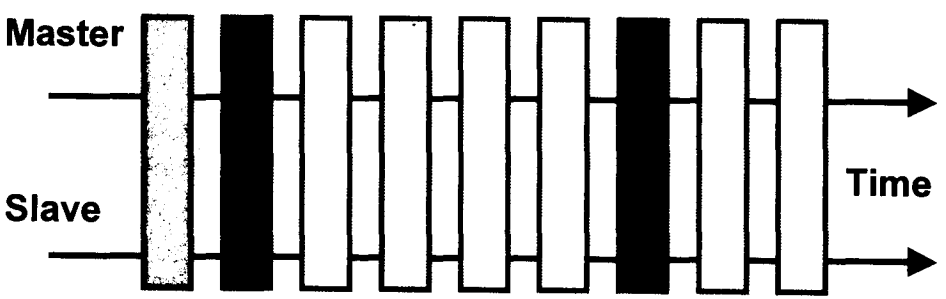
Network Topology of Bluetooth



Support both point-to-point and point –to-multi-point connection

Frequency Hopping in 79 hops displaced by 1 MHz, starting at 2.402 GHz and stopping at 2.480 GHz.



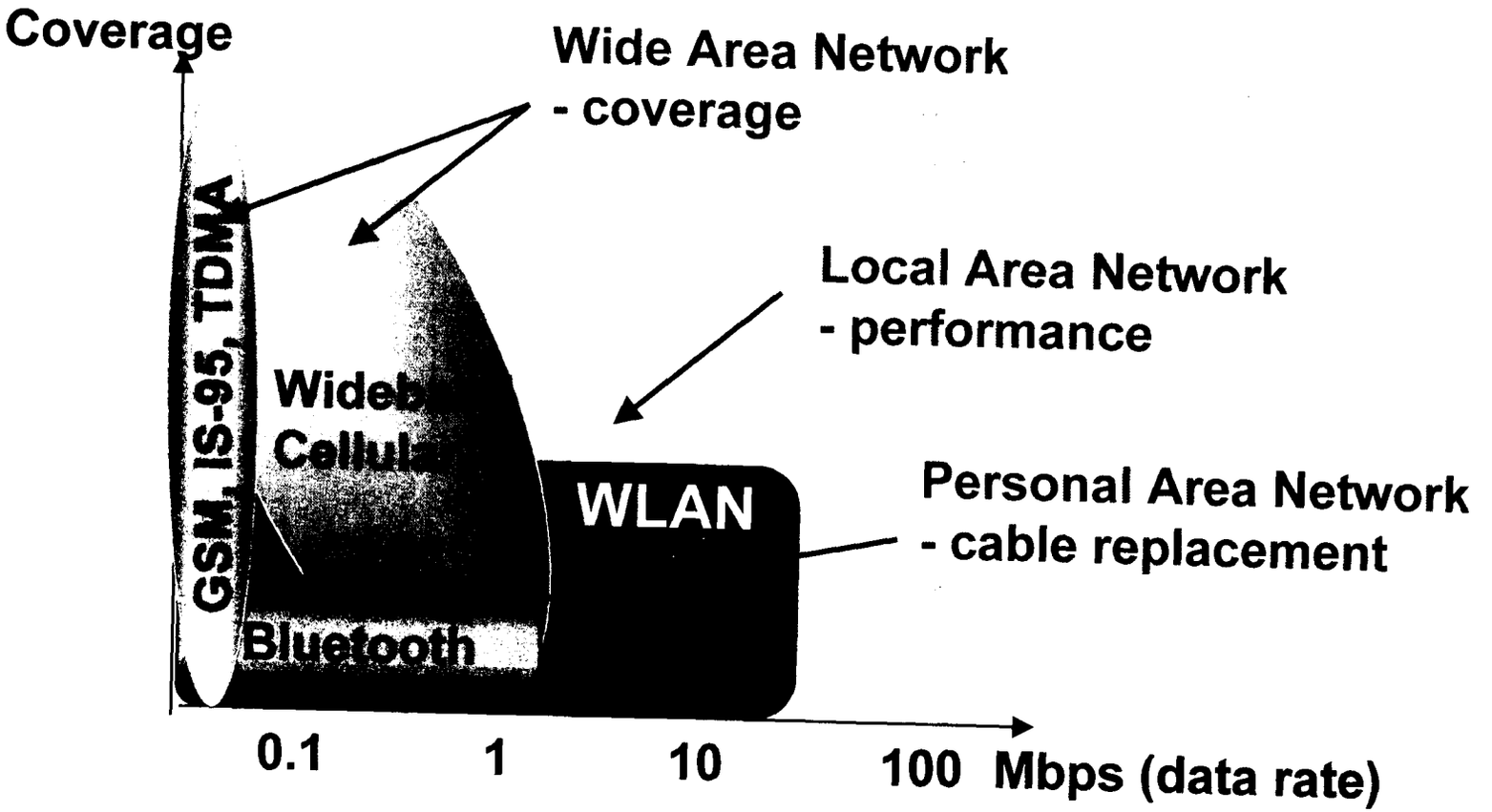


Maximum frequency hopping rate is 1600 hops per second with nominal link range from 10 cm to 10 m.

Confidential and Proprietary

Technology Comparison

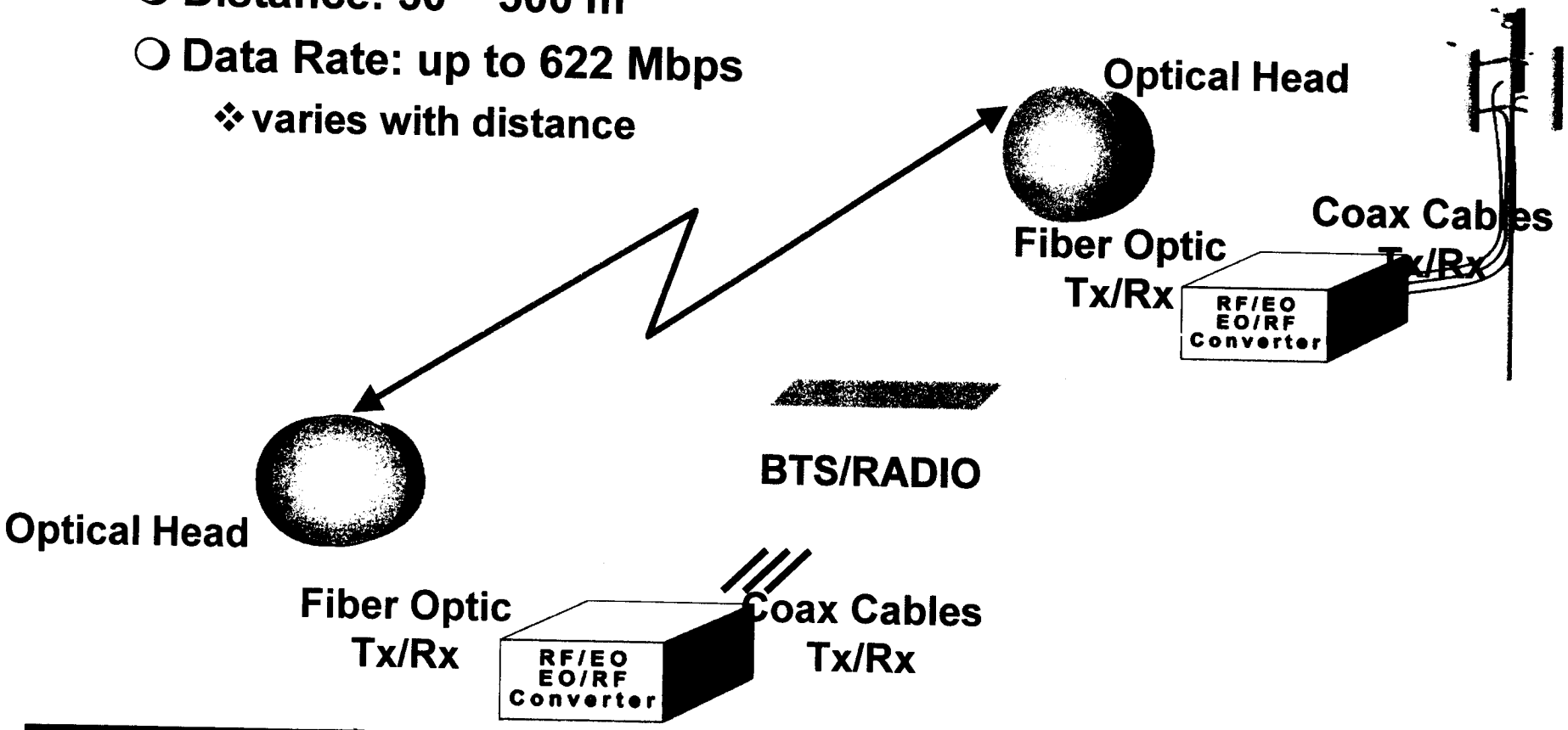
❑ Coverage v.s. Data Rate



Confidential and Proprietary

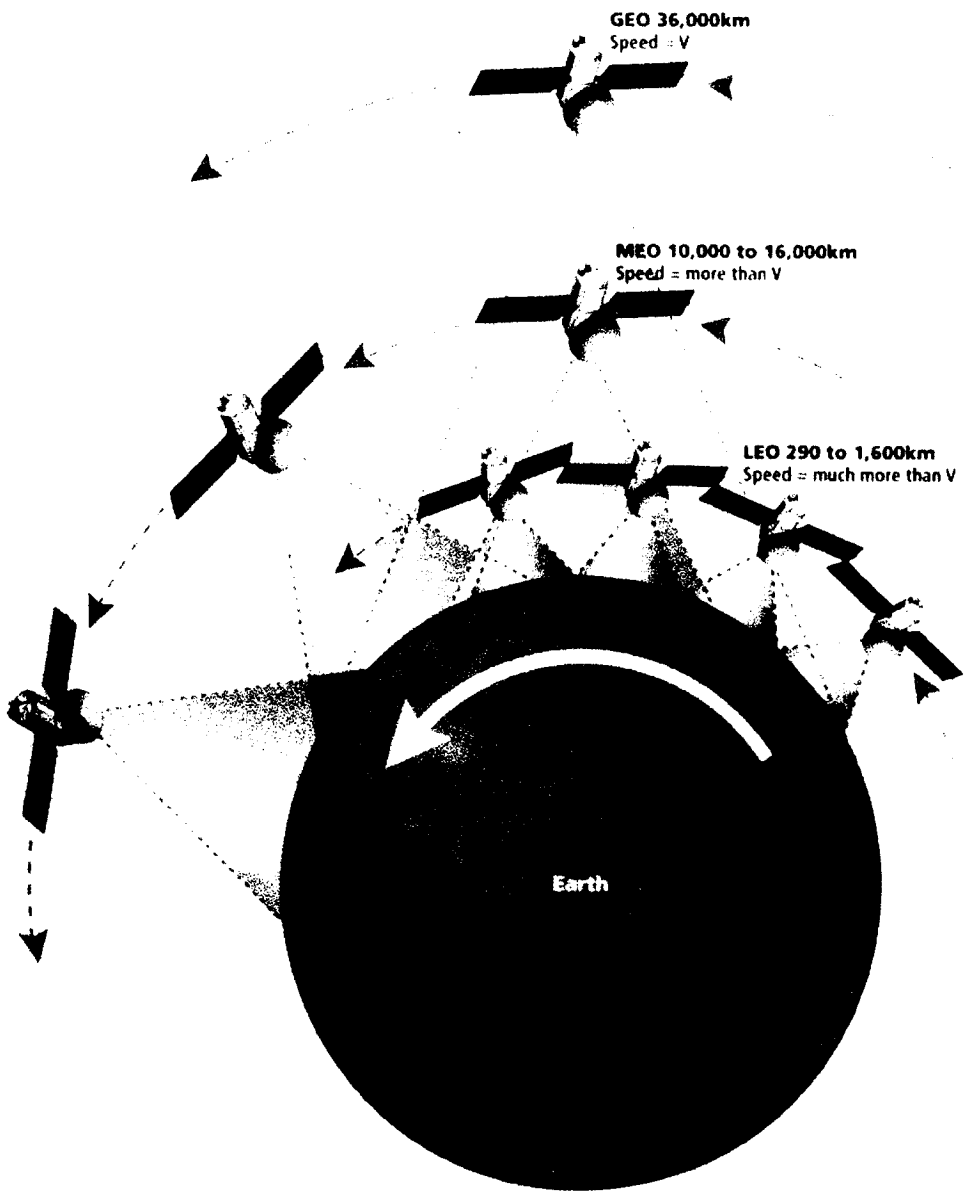
Optical

- Short Distance (PAN) – Infrared
  - Basic component for handset and notebook
  - Support data rate up to 115 Kbps
- Long Distance (WAN) – Free Space Optical (FSO)
  - Distance: 30 ~ 500 m
  - Data Rate: up to 622 Mbps
    - varies with distance



Confidential and Proprietary

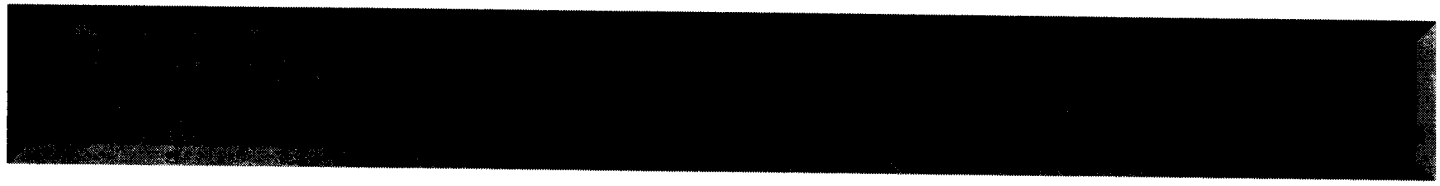
Satellite



	Characteristic
Cost of launch	Highest for GEO due to distance required to achieve orbital position
Number to cover Earth	4 for GEO 12 for MEO 50 for LEO
Power	625 times greater for GEO than for LEO
Atmospheric drag	Greatest for LEO
Ground station tracking	None required for GEO, because the satellite appears to be stationary
Latency	~260 ms for GEO <110 ms for LEO

Confidential and Proprietary

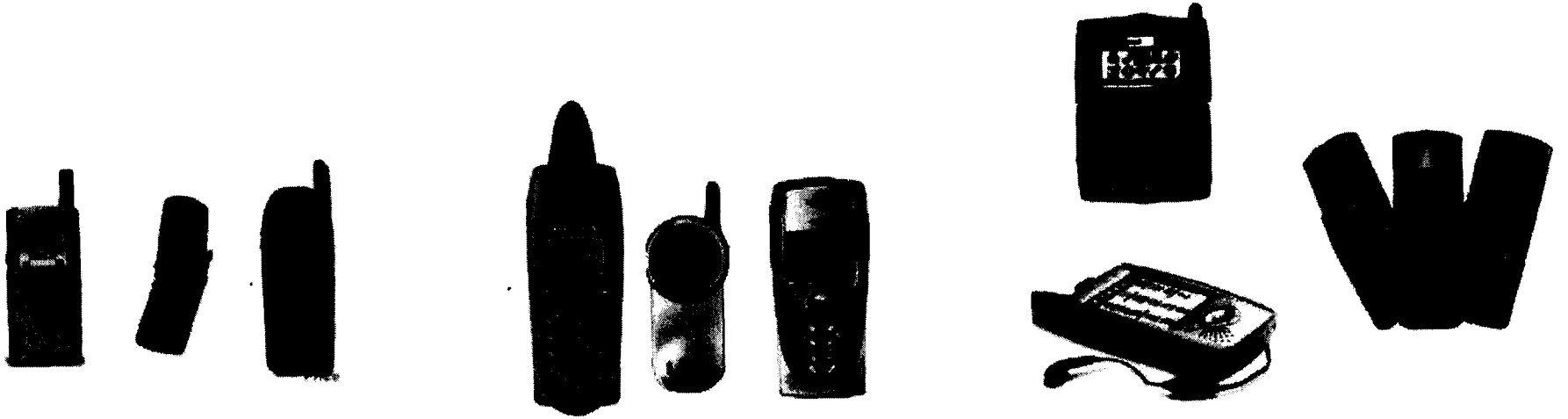
3



Confidential and Proprietary

Evolution of Mobile Service

Voice → SMS WAP → HTML/XML MExE CAMEL



Voice Service
- Voice Call
- Voice Mail

Data Service
- E-mail
- Text Browsing

Personal Service
- Dynamic Download
- Interactive Graphic
- Offline Operation
- Security
- OTA Setting

Confidential and Proprietary

Key Service Areas

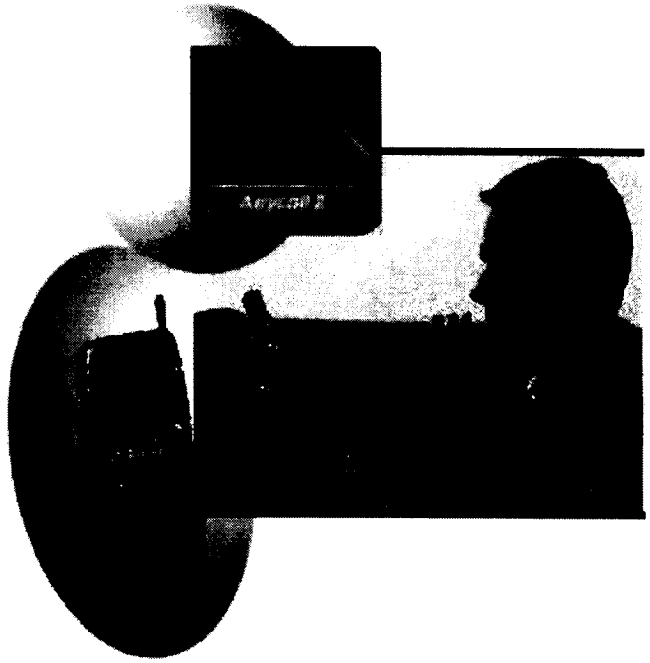
- Multimedia Message (MM)
  - SMS augmented with multimedia content
- Wireless Internet searches
  - Mobility can add value to the search
- LoCation based Services (LCS)
  - Support wireless access
  - Add value to Internet searches
- Entertainment
- Streaming
  - Music, radio, video, news, …
- Talking - augmented to Rich Calls
  - Better sound quality, images
  - Shared view of information while discussing

Confidential and Proprietary

Trend of Wireless Service

- High Data Rate → Technology Improvement
  - Immediately transmission for larger data (information)
- Personalized Information → Content Improvement
  - Anytime, Anywhere, Any-type
- Simple and Easy to Access → User Interface Improvement
  - Integrated and unified





Confidential and Proprietary

Schemas and Ontologies: RDF Schemas

- Ontologies and schemas are ways to describe meaning and relationships of terms
- Define ontology in terms of RDF means RDF schema
- A schema:

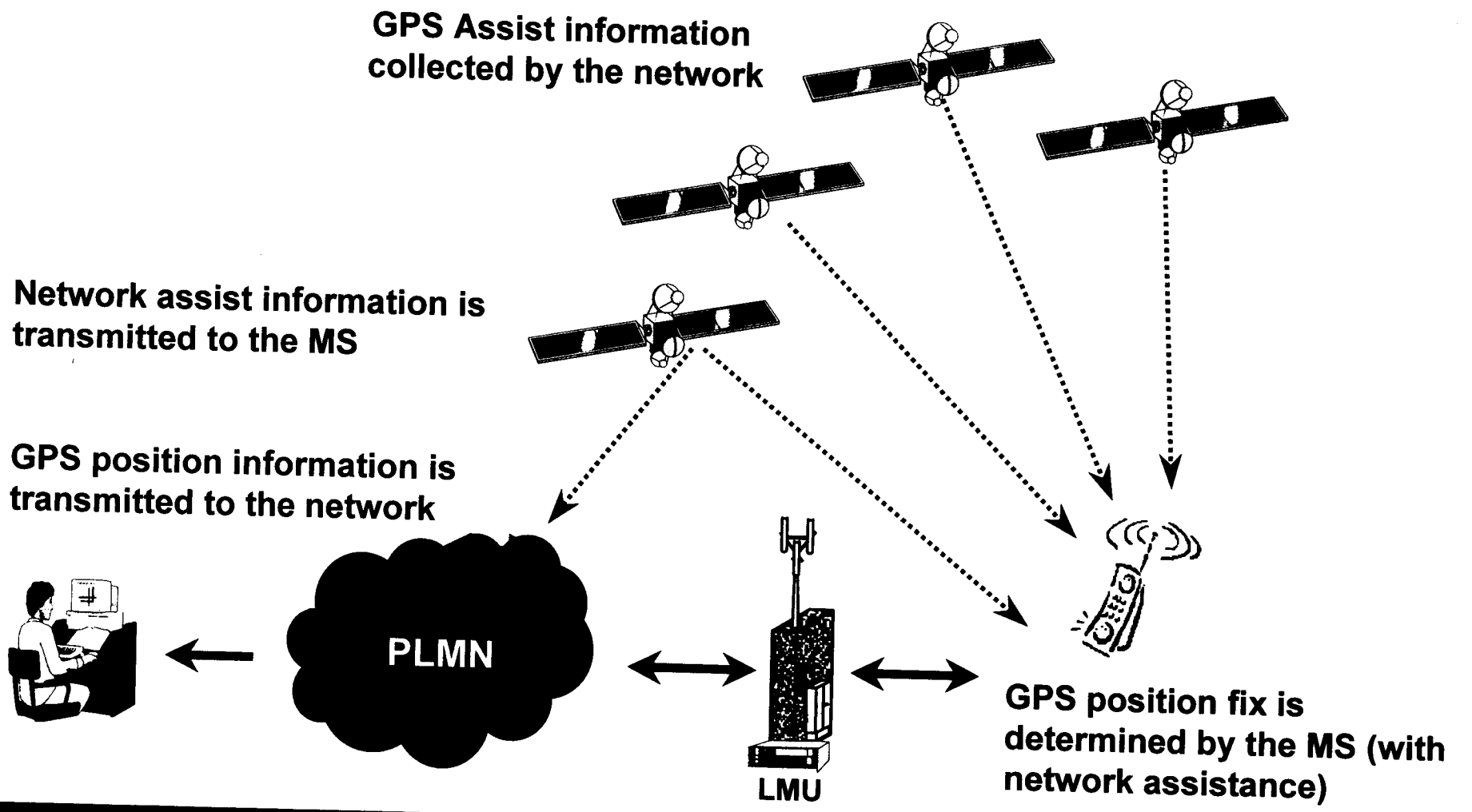
@prefix dc:<http??purl.org/dc/elements/1.1/>
@prefix rdfs: http://www.w3.org/2000/01/rdf-schema#
An author is a type of contributor:
dc:author rdfs:subClassOf dc:contributor

RDF Schema

- Is a set of pre-defined resources and relationships between them that define a simple meta-model including concepts of
  - class,
  - property,
  - subclass and subproperty relationships,
  - domain and range of property constraints
  - and so on.

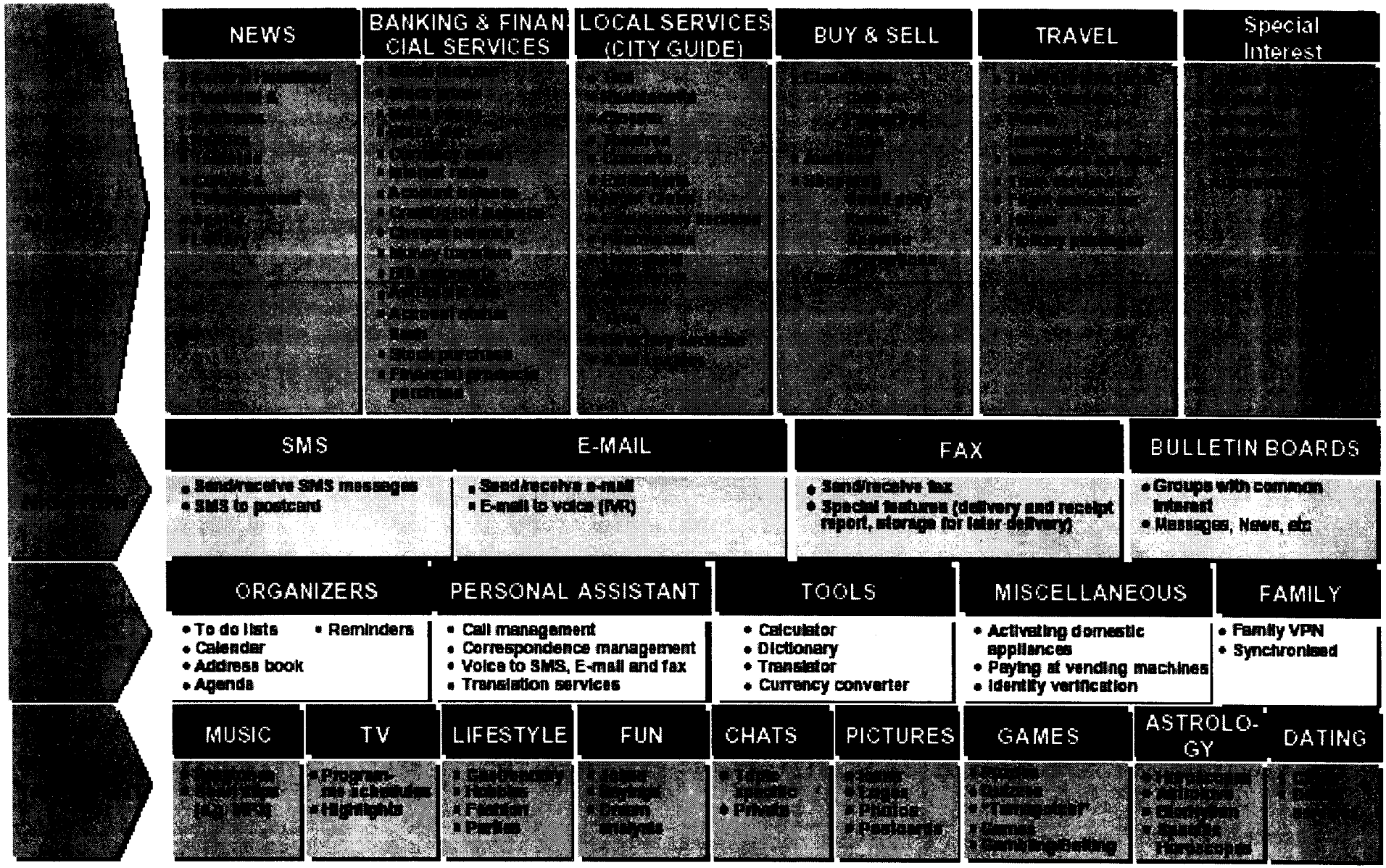
Position Service

Integrate multiple technologies to provide personal location.



Confidential and Proprietary

Application Hypermarket



Source: Nokia

Confidential and Proprietary

4

[redacted]

Confidential and Proprietary

XHTML May Be The Future

- XHTML brings WAP and WWW together
  - XHTML as the basis for WAP 2.0
  - One language for both worlds, style sheets take care of the rest
- Enhanced layout control and functionality across all devices
  - Tables, borders, positioning, aligning, images and image links…
- Consistent appearance on all browsers
- Smaller memory foorprint and enhanced performance efficiency
- Industry-wide support (especially WAP Forum, W3C)
  - Based on HTML tags
  - W3C recommendation for future web development
- cHTML limitations
  - Not a W3C standard
  - Does not support cascading style sheets

Confidential and Proprietary

All IP

- Based on IPv6 in 3G, enhanced rich call control with SIP
- 3G is a natural evolution path towards ALL-IP
- Enables new kinds of services and applications
  - Ubiquitous internet access
  - The services and applications become transparent
  - Proper end-to-end quality of service and security
- Revolutionizes the cost structure of revenues
  - Today voice calls constitute about 95% of revenues
- Different kinds of networks can be connected together
  - Wholly new devices connected to the network
  - Metcalfe effect

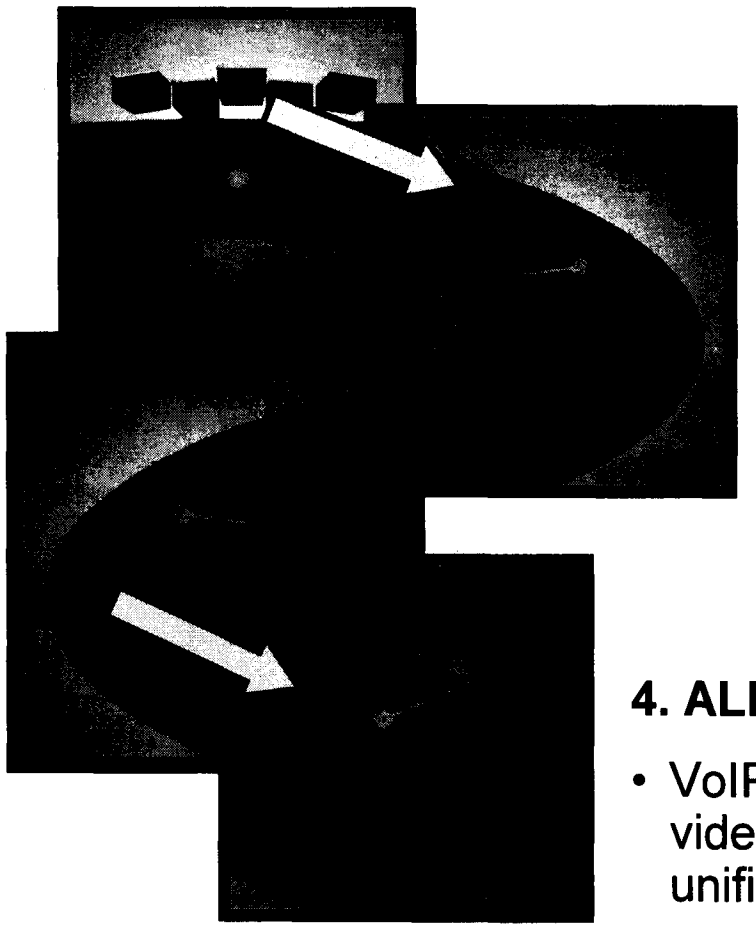
Confidential and Proprietary

Technology Trend Toward All IP

1. 2G (Digital Mobile Com.)

Traditional telecommunication

- Application: Voice
- Technologies: GSM and other cellular protocols



2. GSM data

Appearance of data based services

- New applications: Normal messaging like SMS, WAP, tones & icons
- Technologies mainly as previously + HSCSD.

3. Packet Radio (Rel'97&'99)

Adapting packet switched non real time services

- WAP browsing over GRPS
- Multimedia messaging
- Always on functionality
- Enhanced interoperabilty with IP protocols
- Technologies: GPRS, UMTS

⇨ Evolution

➡ Revolution

4. ALL-IP (3GPP Rel5)

- VoIP over air: IP telephony, video conference, rich call, unified messaging
- Multiple simultaneously active applications
- Technologies: IPv6 on top of several bearers, SIP, GERAN, UTRAN, ...

Confidential and Proprietary

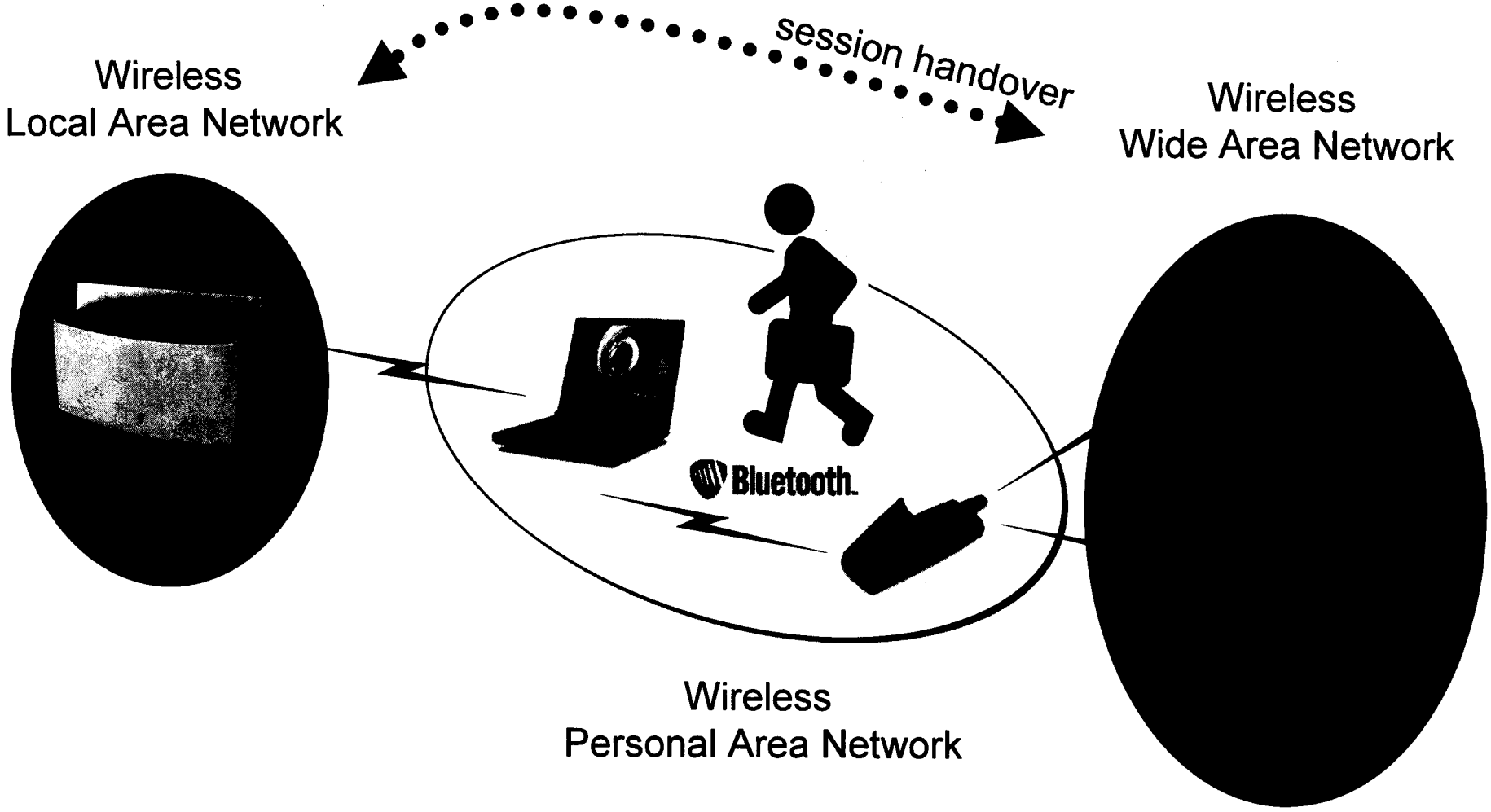
Beyond 3G

- 3G will continue to evolve after initial deployment
  - Evolution towards IP-based core networks with multi-radio access, integration of 3G and WLAN will offer data rates from 10 to 100 Mbps
  - In parallel, a revolution expected to happen within 10-15 years
- Key driver for systems beyond 3G: high data rates everywhere -> hyper-availability of all media
- Unused spectrum does not exist - long term planning necessary to re-farm or re-allocate bands. WRC 2006 expected to consider requirements and identification of spectrum for systems beyond 3G
- New basic research needed, Wireless World research Forum already established
- Vision for revolutionary systems: 100 Mbps in full mobility, 1 Gbps in low mobility environments

Confidential and Proprietary

Technology Convergence

❑ Always Best Connection



Confidential and Proprietary

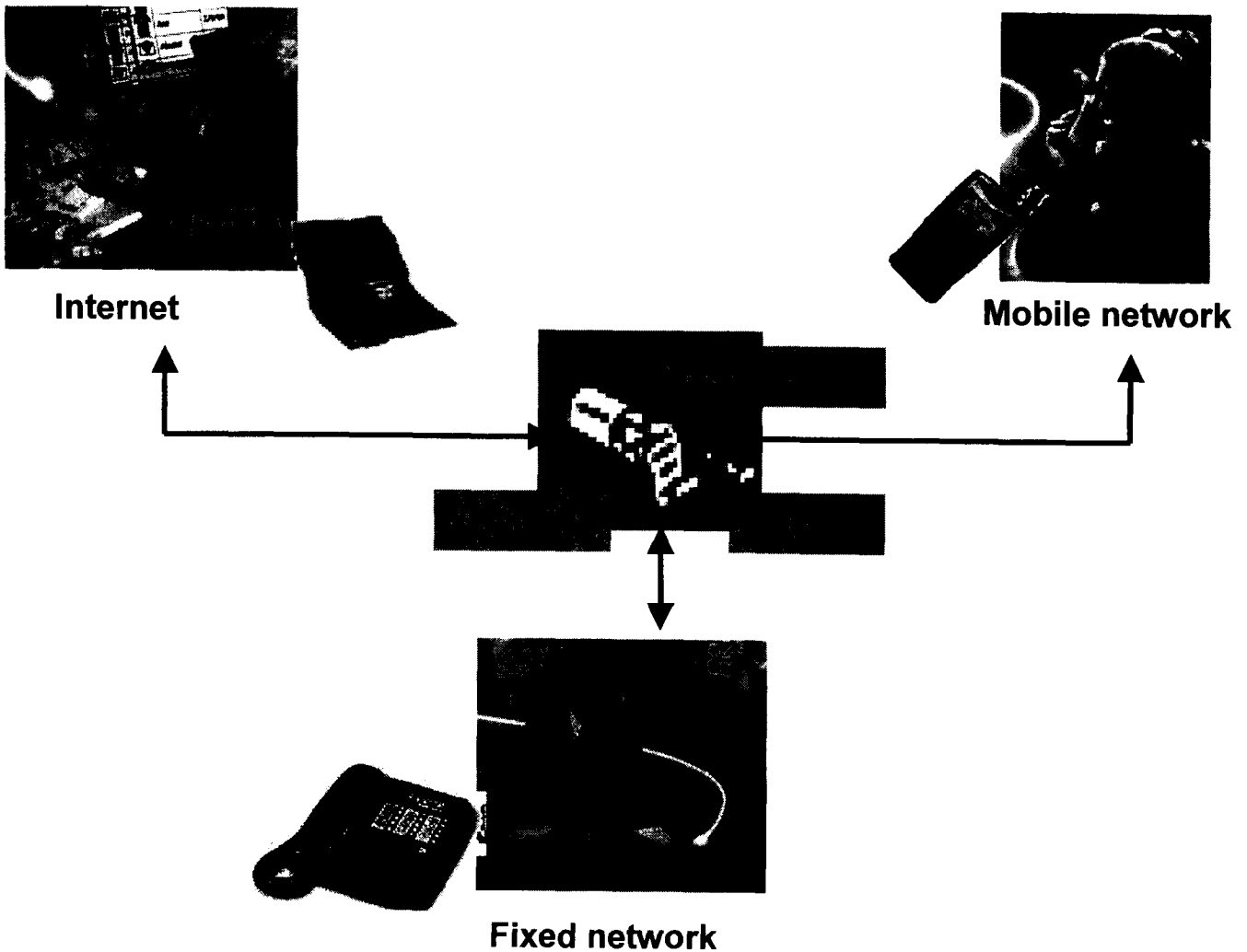
Technology Summary

- Next Generation systems will be defined by services. It will be a mix of low and high-latency applications.
- High data-rate service coverage may not be ubiquitous but small cells on a dense broadband network.
- The network will be unified with an IP backhaul.
- The air-interface may be unified with OFDM.
- Dynamic resource allocation will be a common theme for layers one and two to maximize bps/Hz/$.
- Dynamic network allocation to match resources with traffic density.

Source: 3G World Congress

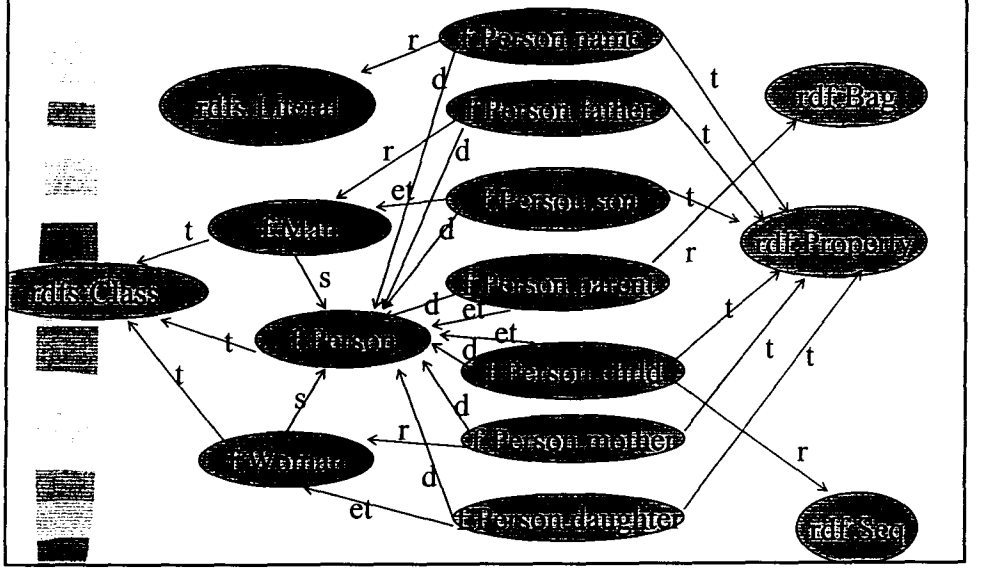
Confidential and Proprietary

Unified Service



Confidential and Proprietary

Family Ontology in terms of RDF schema



Property Labels and Namespace Abbreviations

t = rdf:type

s = rdfs:subClassOf

d = rdfs:domain

r = rdfs:range

et = rdfsx:collectionElementType

rdf = http://www.w3.org/1999/02/22-rdf-syntax-ns#ns#

rdfs = http://www.w3.org/2000/01/rdf-schema#

rdfsx = http://nzdis.otago.ac.nz/0_1/rdf-schema-x#

f = any new namespace chosen for this schema

Unified Billing

❑ One Bill for all kinds of service

❍ Communication, m(mobile, micro)-Commerce, Credit service

Billing for Mobile Internet

Usage Charges

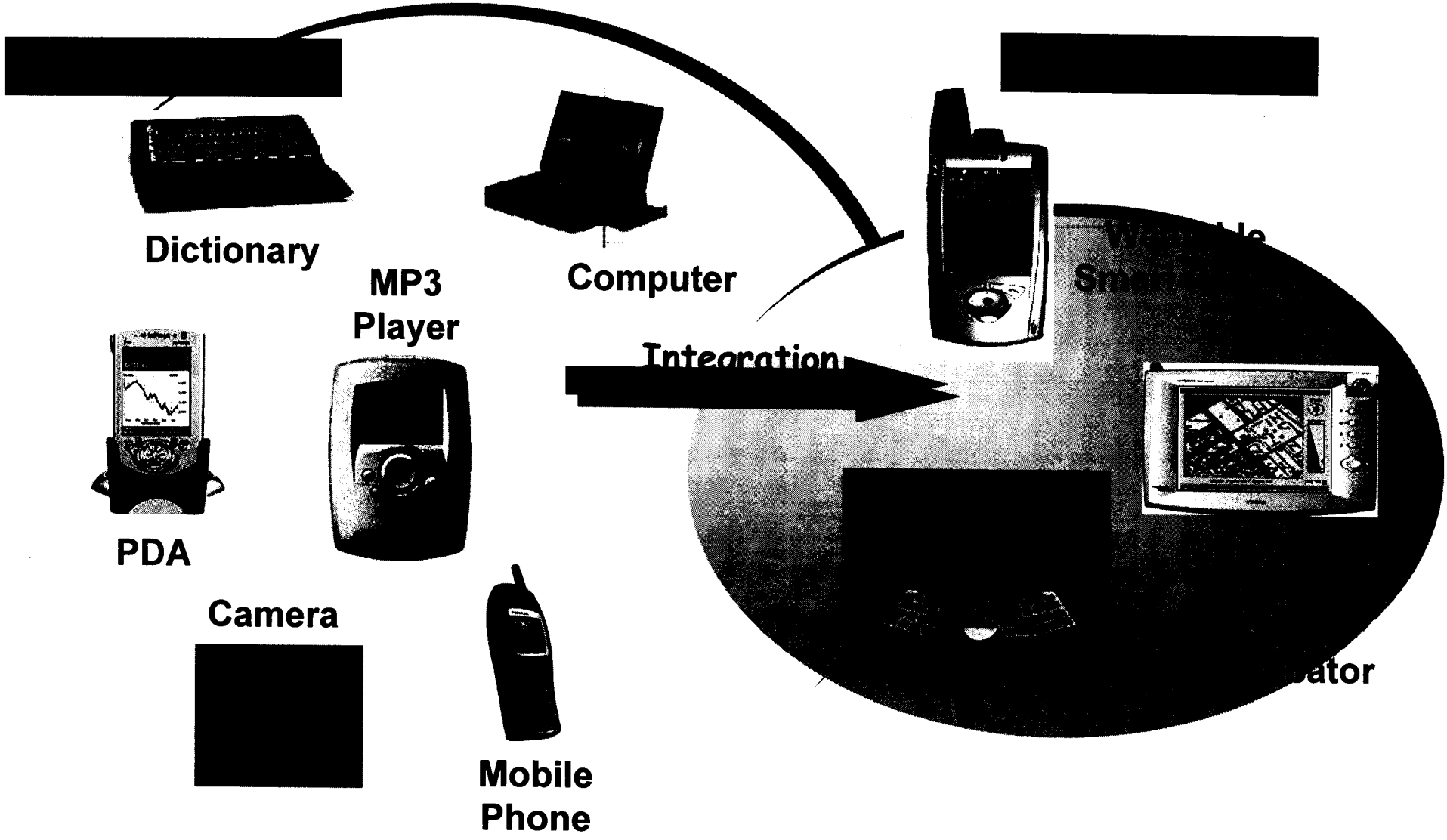
Service	Items	Charge
· Wireless LAN	48	94,78
· GPRS data	58	12,32
· mTicket	13	232,00
	Total:	339,10

Date & Time:	Service:	Information:	Quantity:	Charge:	Currency:
2002-03-08 11:34:32	Wireless LAN	ap001 pphi3g:time	7 h 09 m 07 s	64.3675	EUR
2002-03-08 11:34:32	Wireless LAN	ap001 pphi3g:rcvd	120 kB	0.03	EUR
2002-03-08 11:34:32	Wireless LAN	ap001 pphi3g:sent	219 kB	0.0548	EUR
2002-03-07 09:30:33	Wireless LAN	ap002 pphi3g:time	32 m 34 s	4.885	EUR
2002-03-07 09:30:33	Wireless LAN	ap002 pphi3g:rcvd	25 kB	0.0063	EUR
2002-03-07 09:30:33	Wireless LAN	ap002 pphi3g:sent	152 kB	0.038	EUR
2002-03-04 18:50:15	Wireless LAN	ap001 ppvpn:time	3 m 18 s	0.495	EUR
2002-03-04 18:50:15	Wireless LAN	ap001 ppvpn:rcvd	0 Bytes	0.0003	EUR
2002-03-04 18:50:15	Wireless LAN	ap001 ppvpn:sent	814 kB	0.2035	EUR
2002-03-04 13:33:26	Wireless LAN	ap001 ppvpn:time	6 m 48 s	1.02	EUR
2002-03-04 13:33:26	Wireless LAN	ap001 ppvpn:rcvd	0 Bytes	0.0003	EUR

Confidential and Proprietary

Terminals

- Portability: key factor for user to carry



Confidential and Proprietary

SW of Terminals

- Unlike hardware integration, SW will be competitive diverse



Symbian EPOC



Microsoft WinCE

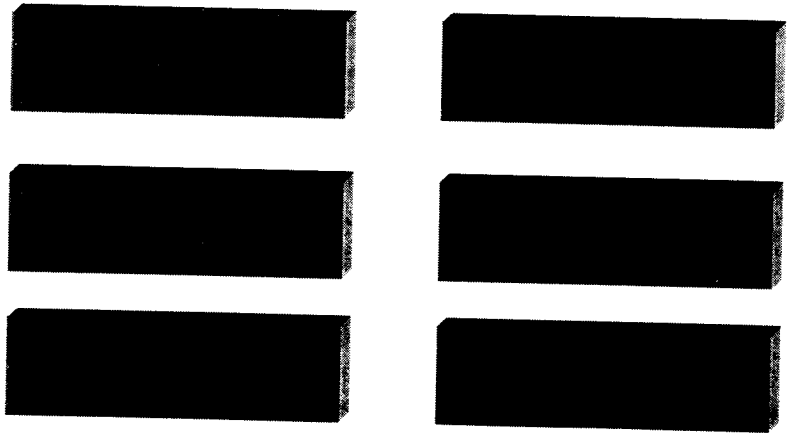


Palm OS



Java

MExE Classmark III



Confidential and Proprietary

5

Confidential and Proprietary

- Wireless module will be the basic function of end terminals
  - IR, Mobile, Bluetooth, WLAN, GPS
- Most terminals will have similarity in functions
- Segmentation and Integration of wireless PAN/LAN/WAN will be the trend
- TCP/IP is the basic function
- New Value is coming from the Content of Data Service
- User Friendly and Attractive Figure are the competitive points for end terminals
  - OTA

Confidential and Proprietary

Anxiety

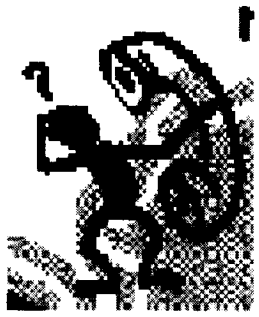
- Wireless is the trend, but what is the business model
  - Bad business model of Fixed Internet will limit the development of Wireless Internet

What kind of customers can I capture?



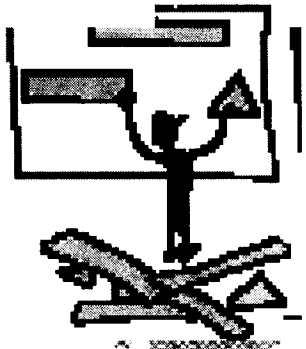
Earn Money ??? or Burn Money !!!





How do I handle customers requests?

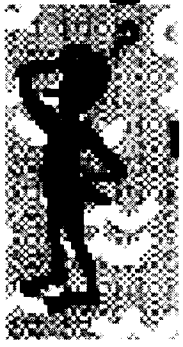
Which new services shall I provide?

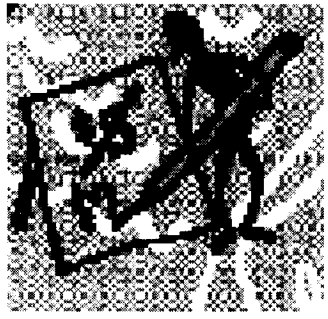




How do I keep my customers?

How do I deliver these new services?





How do I get paid?

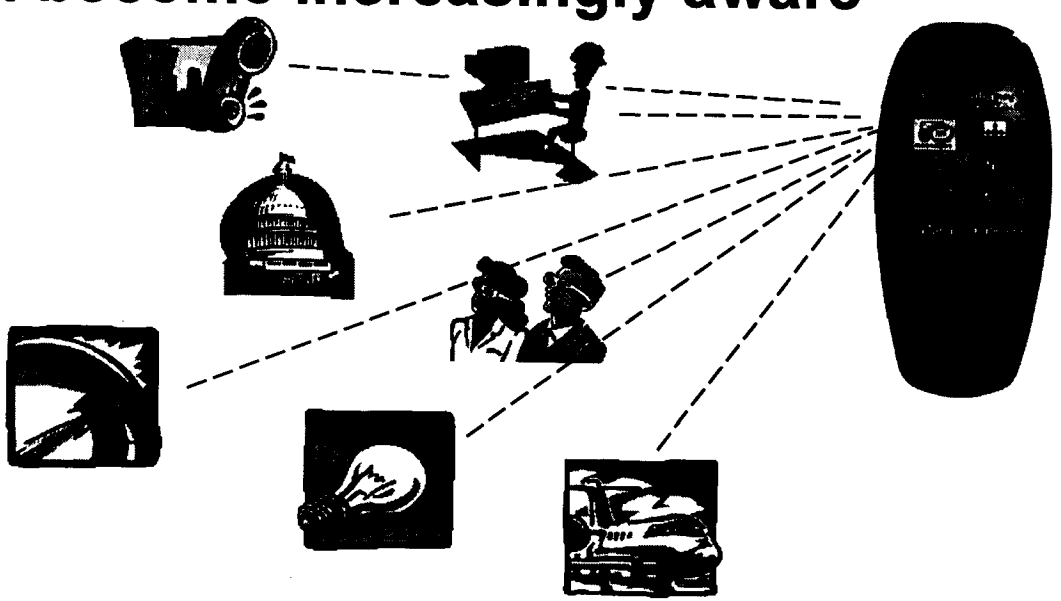


How do I earn more money?

Confidential and Proprietary

Computing/Internet Become Ubiquitous

- Moore's law continues - processing power today at insect brain level, mouse brain in 2010, human brain in 2020
- In 2003, a typical home will contain 380 microprocessors
- IPv6: 128 bit address - address space 3.4×10^{38}
  - 3×10^{24} IP addresses per each m^2 on Earth
  - Totally as many as the estimated number of sand grains on Earth ...
- The personal terminal will become increasingly aware about its environment:
  - People
  - Places
  - Networks
  - Services
  - Other devices
  - ...



Sources: R. Kurtzweil: The Age of Spiritual Machines; Dataquest; Prosessor

Confidential and Proprietary

Q & A

Thank You!!!

Confidential and Proprietary

工業技術研究院
電腦與通訊工業研究所
Industrial Technology Research Institute
Computer & Communications Research Laboratories

Mobile Web Services Workshop

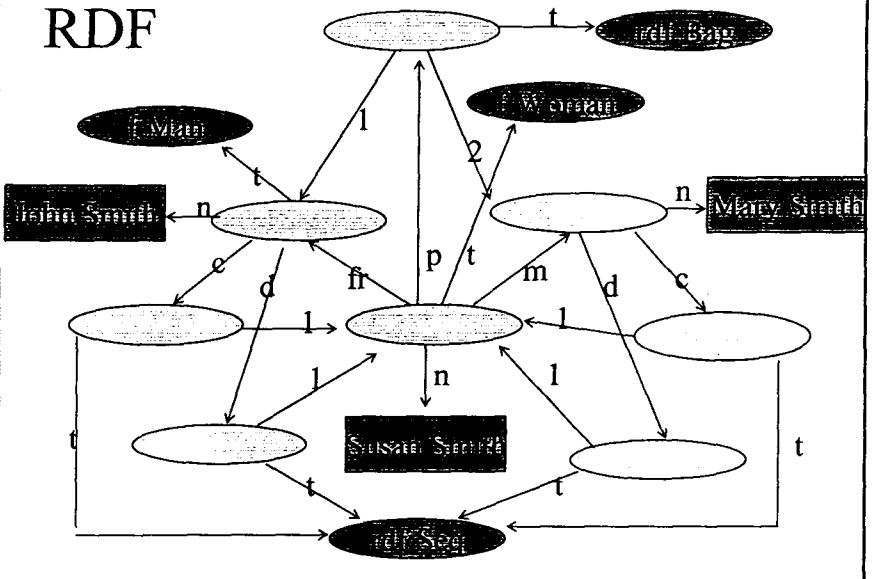
行動式網際網路服務平臺

工研院電通所網際網路軟體技術組
鄭仁傑 Roger Jen-Chieh Cheng
rogercheng@itri.org.tw

Agenda

- ✓ Evolution (What hava happened?)
  - Web Application
  - Mobile Internet
  - Mobile Commerce Area and Market
- ✓ Mobile Service Framework
  - Open Service Access (OSA)
  - MExE
- ✓ Web Services
  - Architecture and Technology
  - P2P Service Gateway
- ✓ Security Issue
- ✓ Conclusion

Family knowledge in terms of RDF



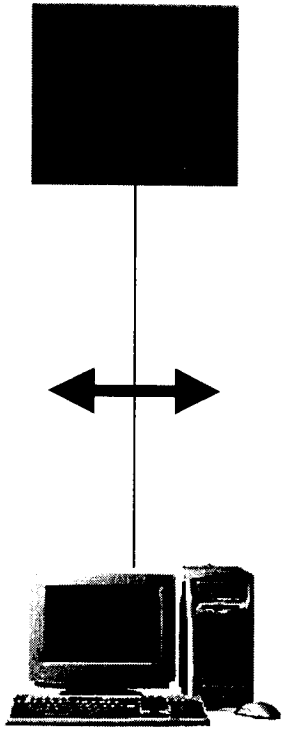
Property Labels and Namespace Abbreviations

t = rdf:type
1 = rdf:_1
2 = rdf:_2
n = f:Person.name
fr = f:Person.father
s = f:Person.son
p = f:Person.parent
e = f:Person.child
m = f:Person.mother
d = f:Person.daughter

rdf = http://www.w3.org/1999/02/22-rdf-syntax-ns#ns#

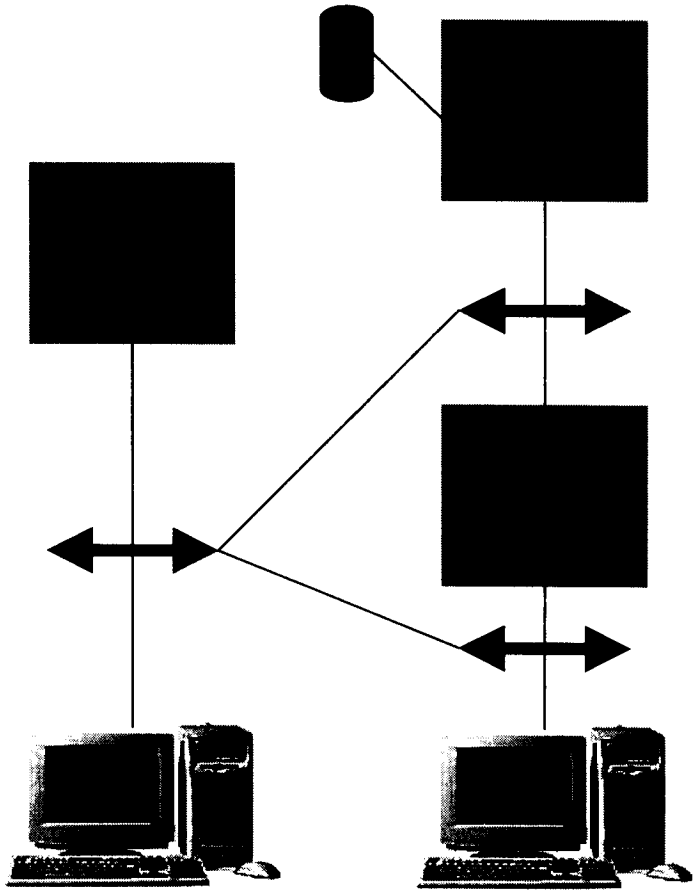
f = namespace chosen in previous rdf schema

Evolution of a Web System - I



Yesterday

Evolution of a Web System - II



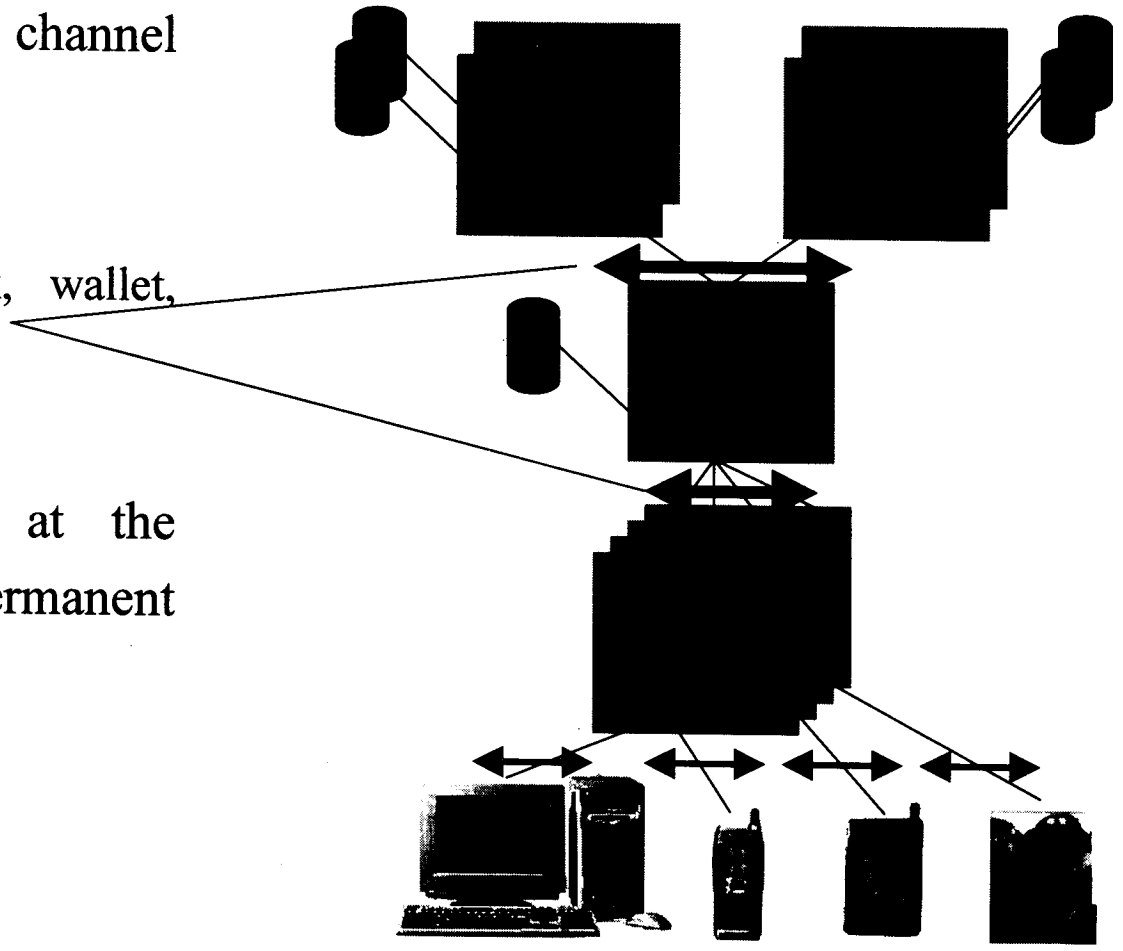
Yesterday Today

- Independent of development between the business logic and the rendering
- More dynamic content as a aggregation of the business logic and rendering operations
  - Profiling and advertizement added at the rendering level
- Scalable solution
- Invisible revolution for the end users

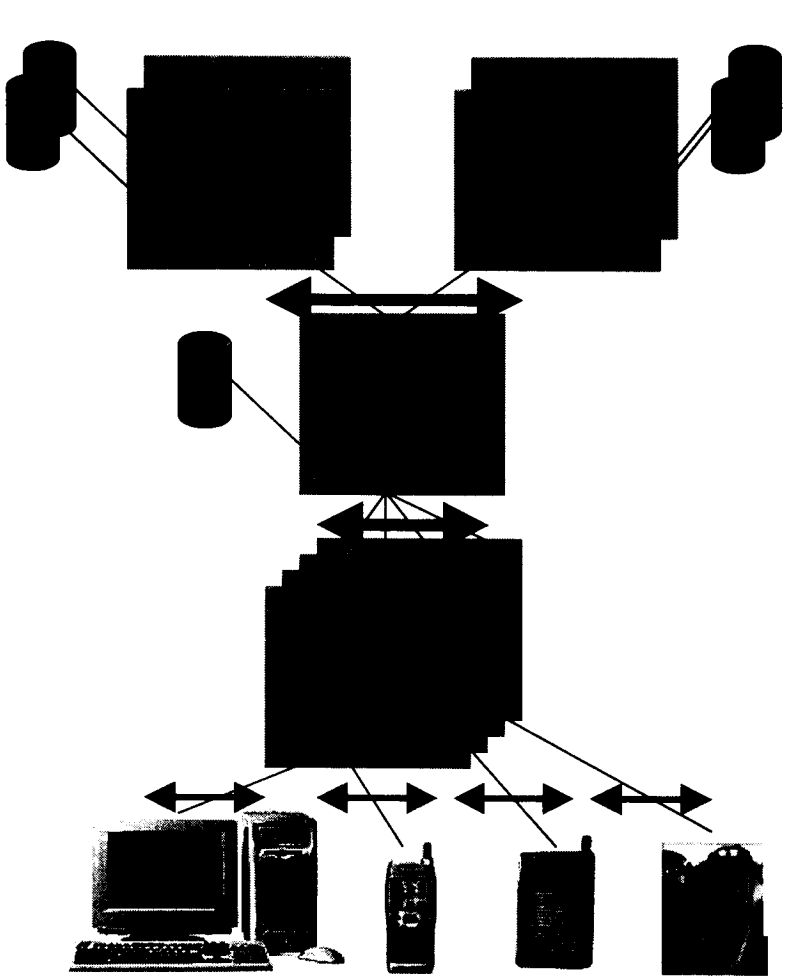
Evolution of a Web System - III

- Multi channel to cross channel transaction
- User based services
  - Calender, address book, wallet, billing…
  - Experience profilling,
- Projection of the user at the network level for permanent presence
- New business creation:

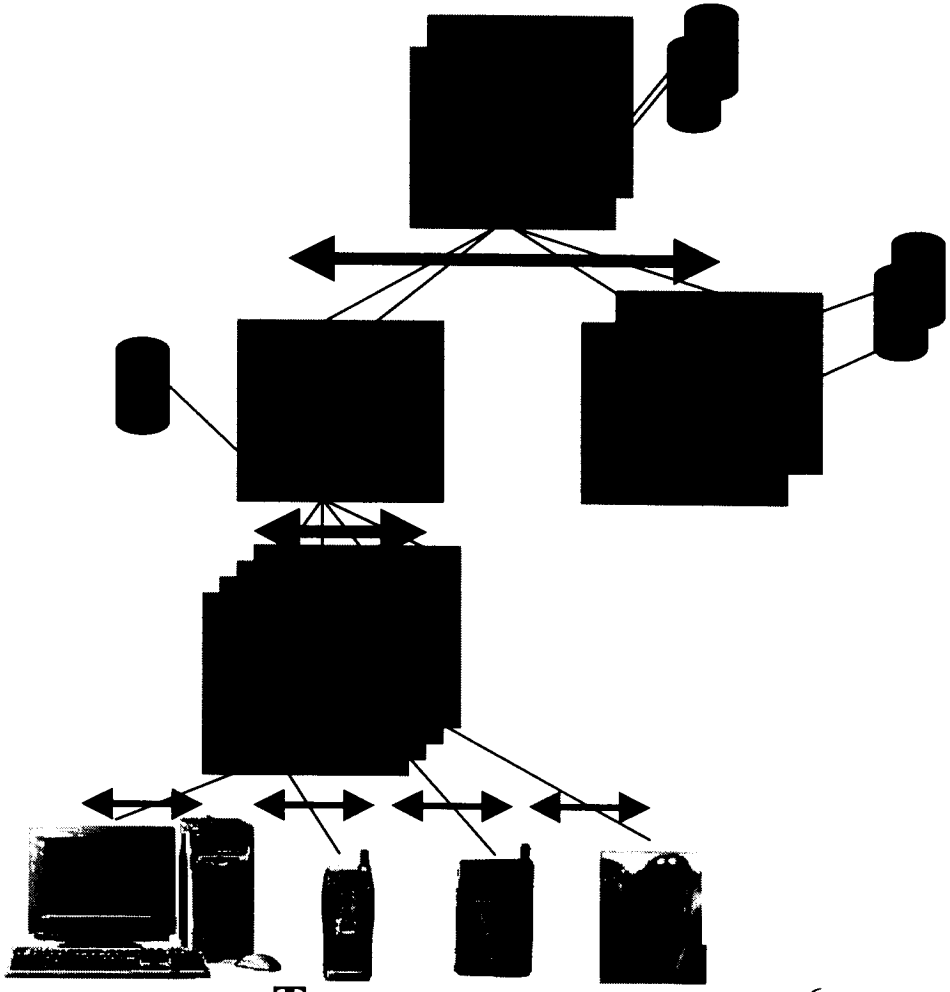
  User Service Provider



Evolution of a Web System - IV



Emerging



Tomorrow

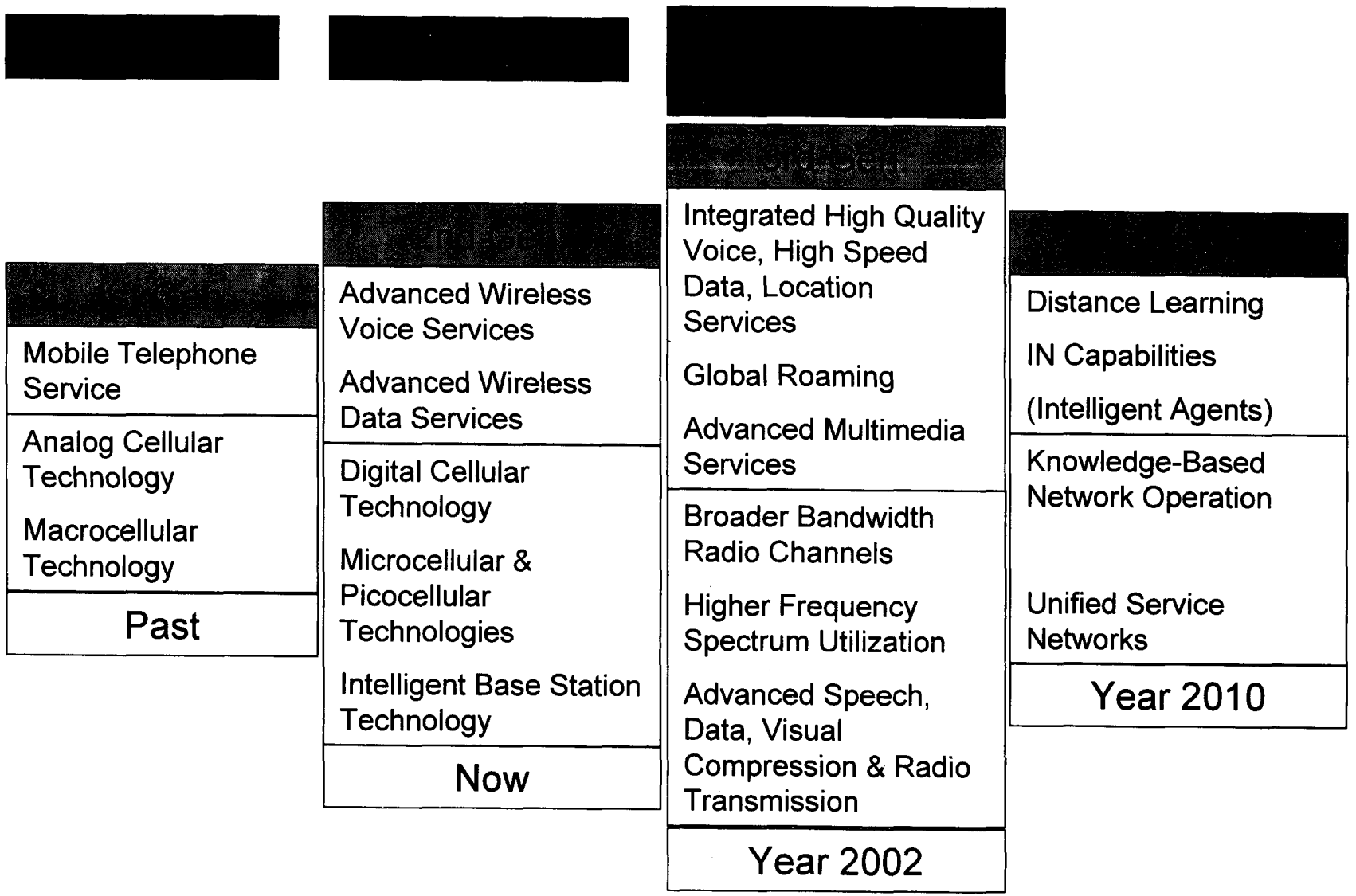
Wireless Today

- Wireless subscribers are demanding new services:
  - Fast internet access
  - Global roaming
  - Wireless commerce
  - Video communication
- Need higher transmission rate and more function features

Wireless Evolution

- 1st generation: 1980s
  - –Analog
  - –Voice Service
- 2nd generation: from 1992
  - –Digital
  - –Voice and low rate data service
- 3rd generation: 2002
  - –High rate data and multimedia service

Wireless Evolution

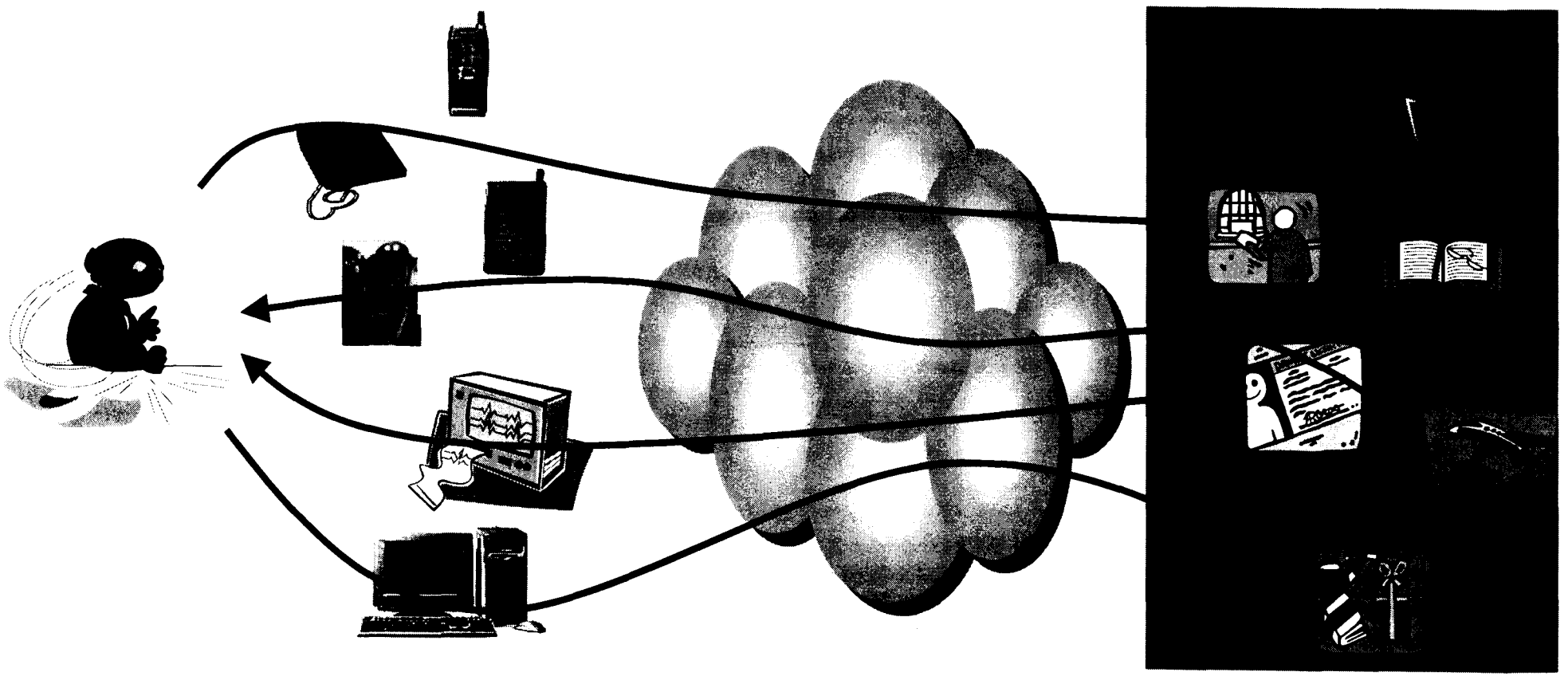


Goals of Network Evolving

- Essence of network evolution: **Convergence of**
  - –communication and distributed computing
  - –Voice, video and data
  - –Telecom and datacom
  - –Control and management
- A ubiquitous QoS-enabled packet-oriented multimedia service network with an open service architecture is the trend.

Today Access is Channel Specific

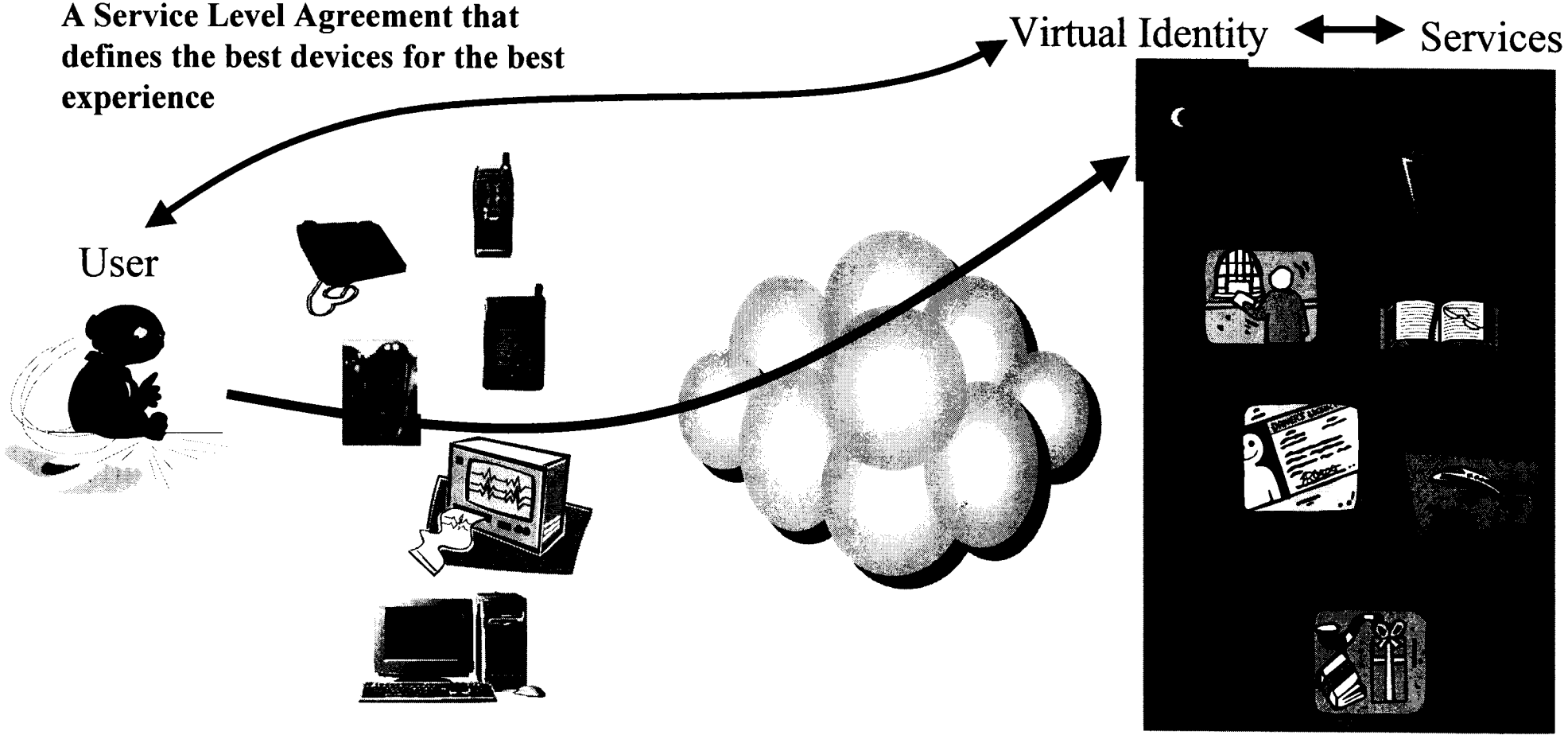
User in context Devices Networks Services



For Each User Experience *For Each Service*

Tomorrow Access is in Two Steps

A Service Level Agreement that defines the best devices for the best experience



Using Sharable Ontology to Retrieve Historical Images



Motivation

- Users might not have the complete historical knowledge for a query. Need the historical ontology.
- For example:
  - I want the picture of Qin dynasty's emperor.
- Our Goal:
  - Establish an image retrieval model with the high precision and easy usage by applying the sharable domain ontology, knowledge and thesaurus.
- The endeavor of semantic web allows domain knowledge to be represented in an interoperable and sharable manner.

Mobile M-Commerce Area

Delivering business / individual's information services
Over Mobile!!!



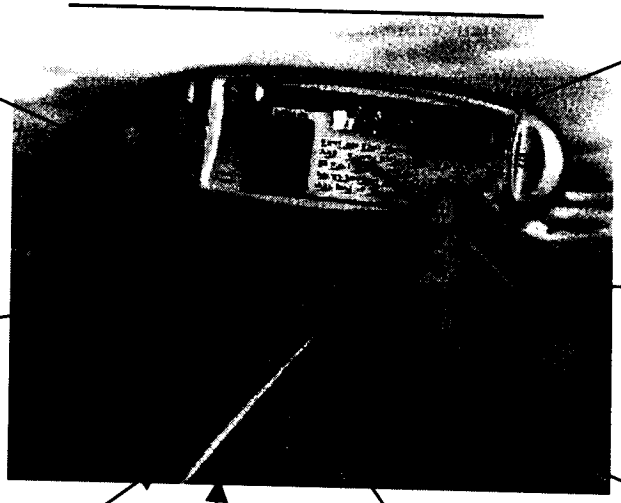


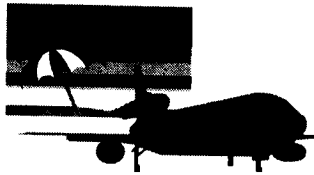
- **Banking**
- **Transacting**



- **Informing**
- **Portal**

News
Sport
Weather
Shares...

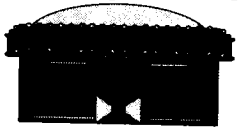




- **Travelling**
- **Scheduling**



- **Video conferencing**



- **Shopping**
- **Entertainment**

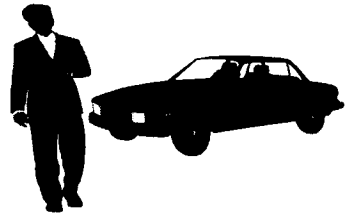


- **New Applications & Services**



- **Database**

- **Positioning**
- **Location**



Mobile E-Commerce (M-Commerce)

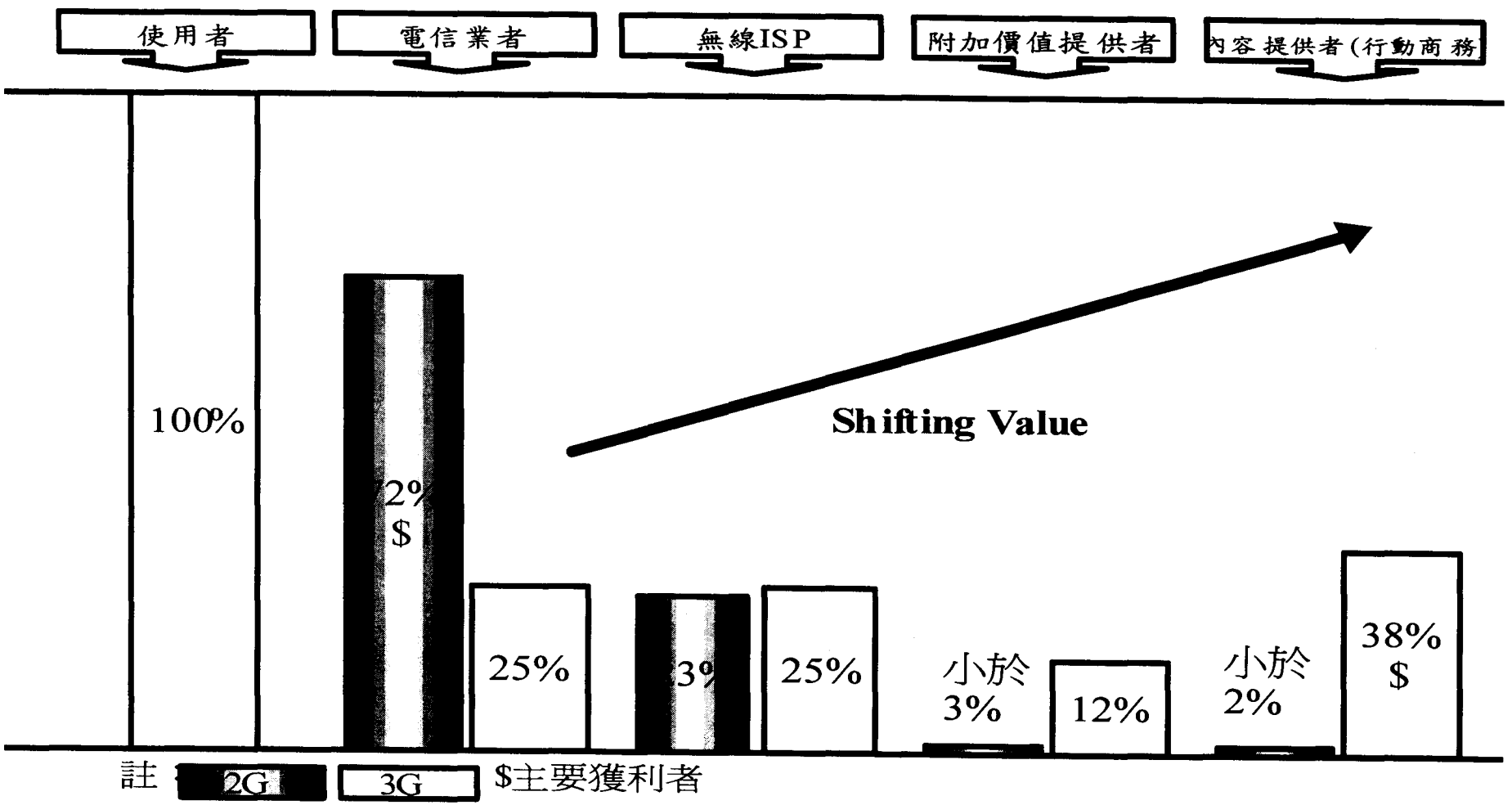
Where can M-Commerce be applied to?

- *B-C (Business to Consumer)*
  - *Mobile banking, Mobile retailing, Mobile stockbroking, Mobile auction, Mobile ticketing, Mobile advertising, Short Message Service, …*
- *B-E (Business to Employee)*
  - *Mobile scheduling, Mobile PIM, Mobile Email, Mobile Learning, Mobile intranet access, Document Retrieval and Management…*
- *B-B (Business to Business)*
  - *Mobile Application Service Provider, Mobile CRM, Mobile ERP, industry-specific application*

Source ：IEK，2001/7

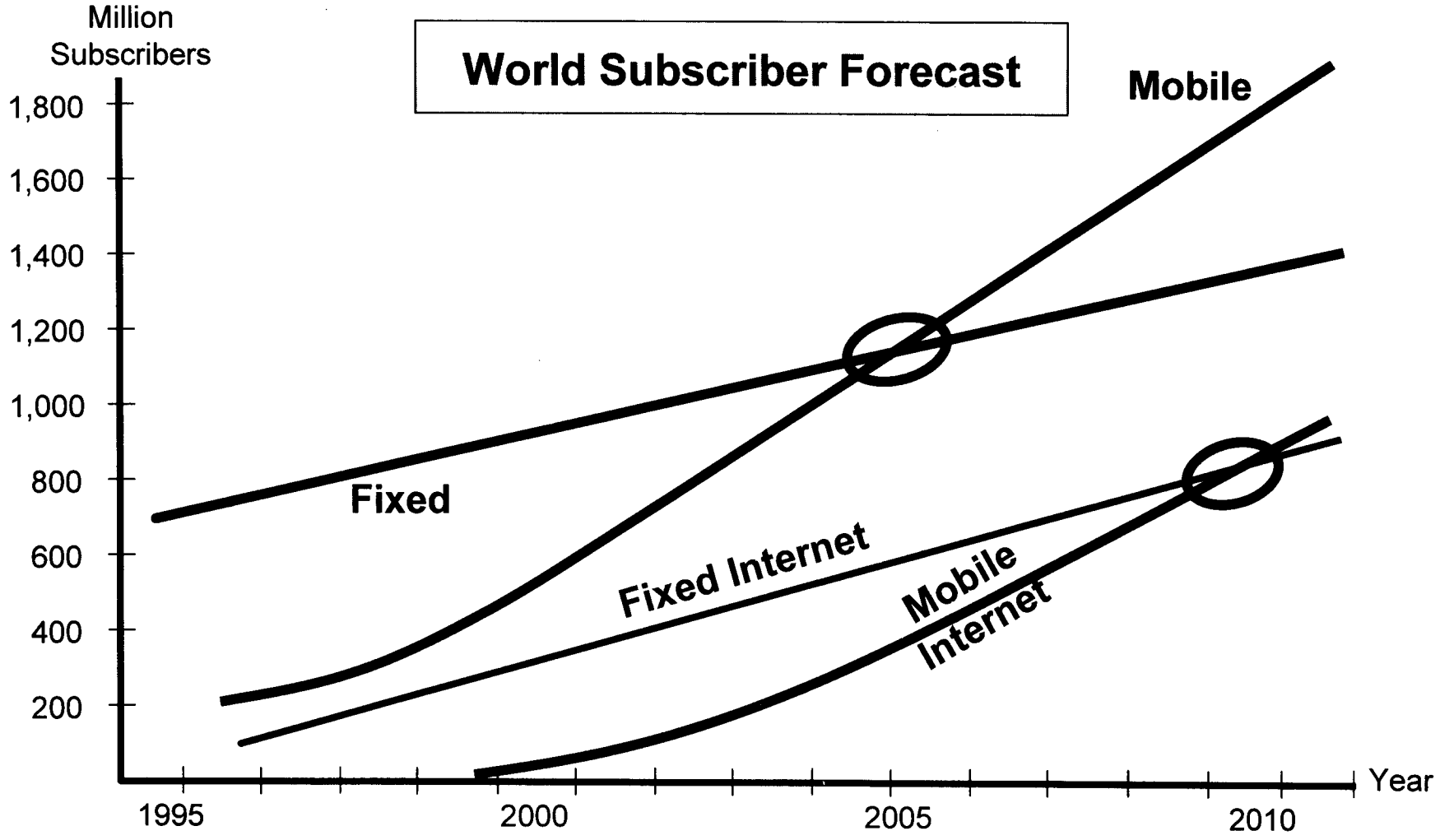
工業技術研究院
電腦與通訊工業研究所
Industrial Technology Research Institute
Computer & Communications Research Laboratories

2G & 3G Mobile Internet Shifting Value



Mobile Market Worldwide

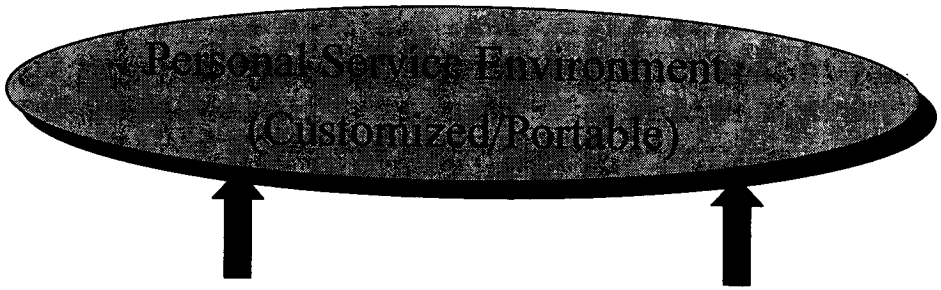
Quick Growth: 1 Billion Internet (mobile & fixed) and mobile subscribers respectively in 2004



Source: Siemens

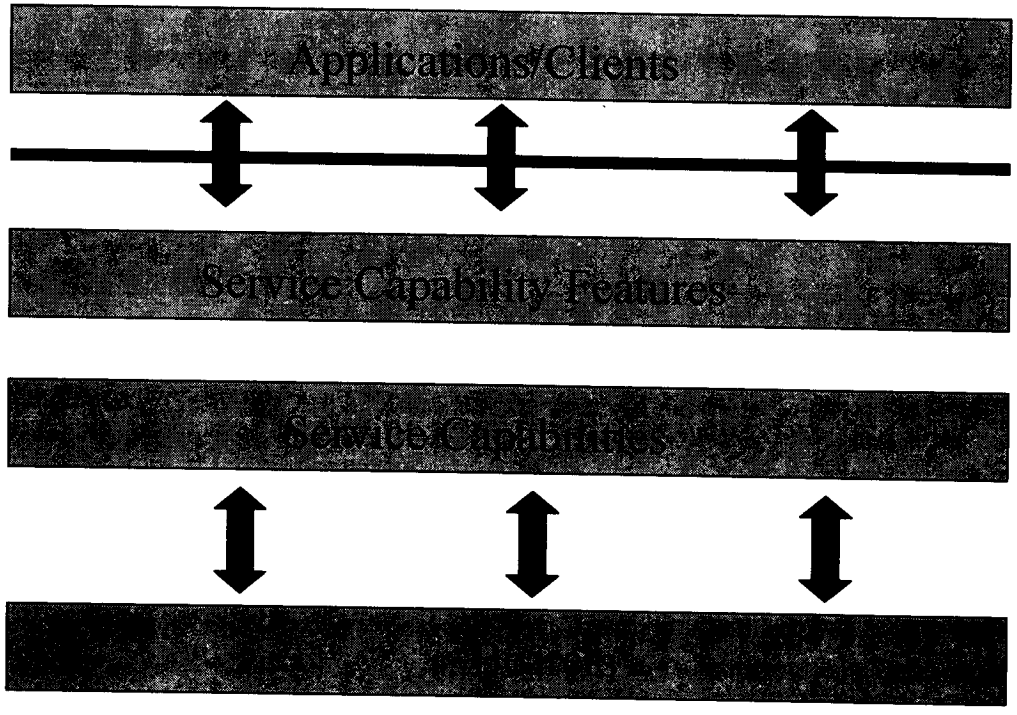
Framework for Services

VHE



❖Personalized services
❖Personalized user interface
❖Consistent set of services anytime and anywhere

OSA



Built using Service Capability Features

Standard API

Pre-set by Standards, e.g. CAMEL, SAT, MExE

Virtual Home Environment (VHE)

- Users are consistently presented with the same
  - personalized features,
  - User interface customization,
  - Services

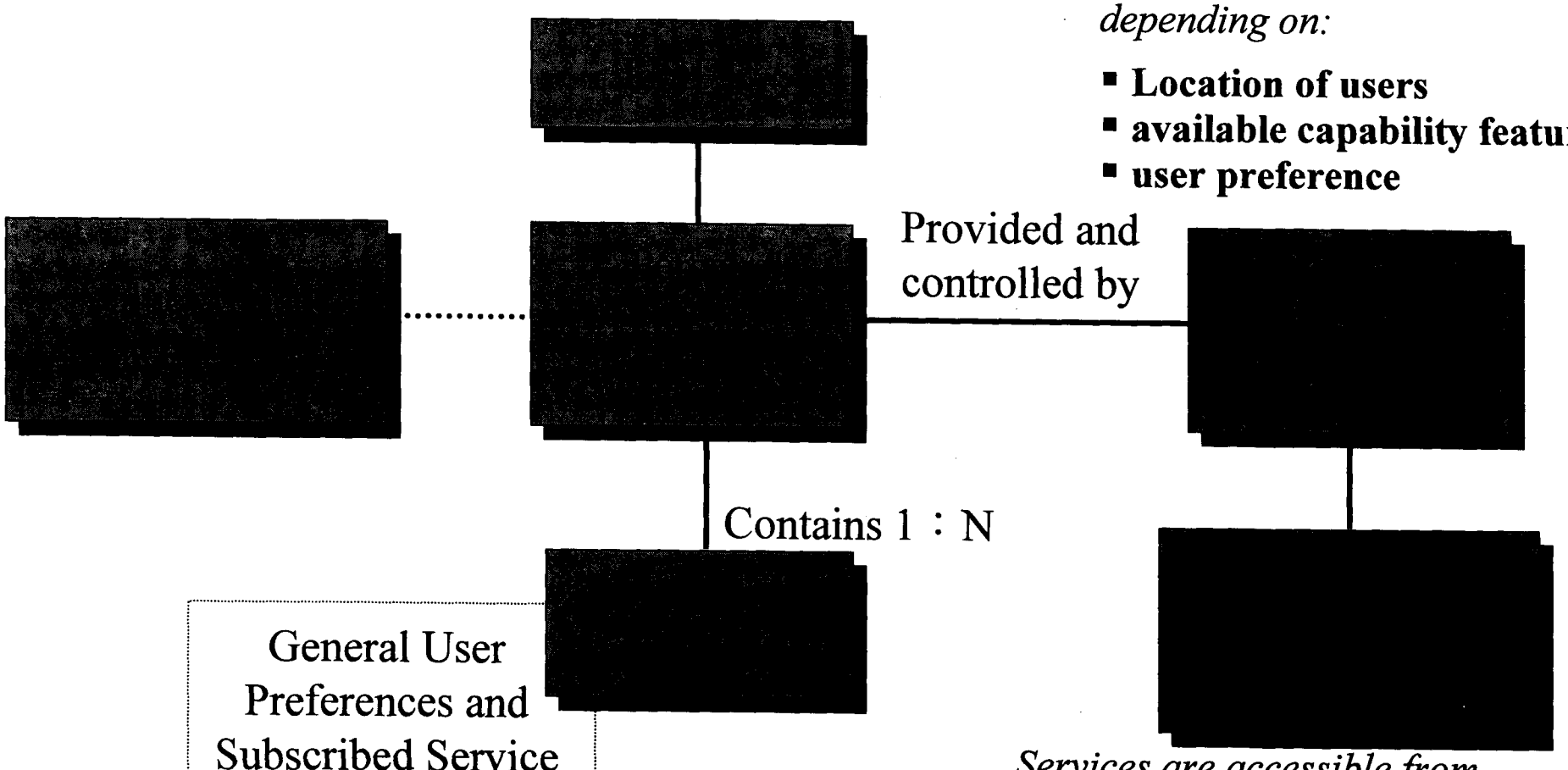
  } → User profiles

  In
  - whatever network→services accessible by all kinds of netowrk
  - and wahtever terminal(within the capabilities of the terminal and network)
  - wherever the user may be located→global roaming
- MExE and OSA are two of the toolkits which may be used to support VHE

VHE

Control Access to services depending on:

- **Location of users**
- **available capability features**
- **user preference**



Services are accessible from

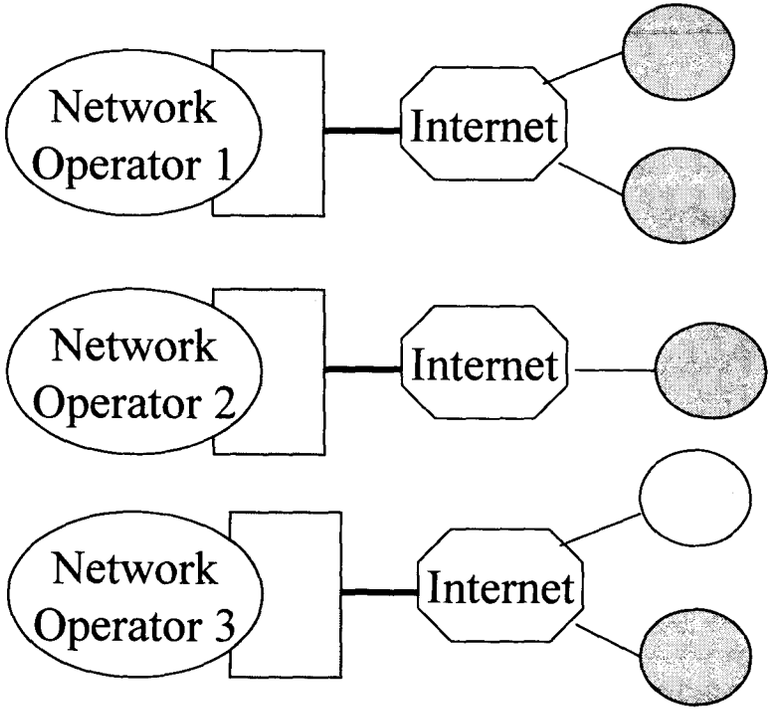
- **Whatever network**
- **wherever users may be**

Open Service Access (OSA)

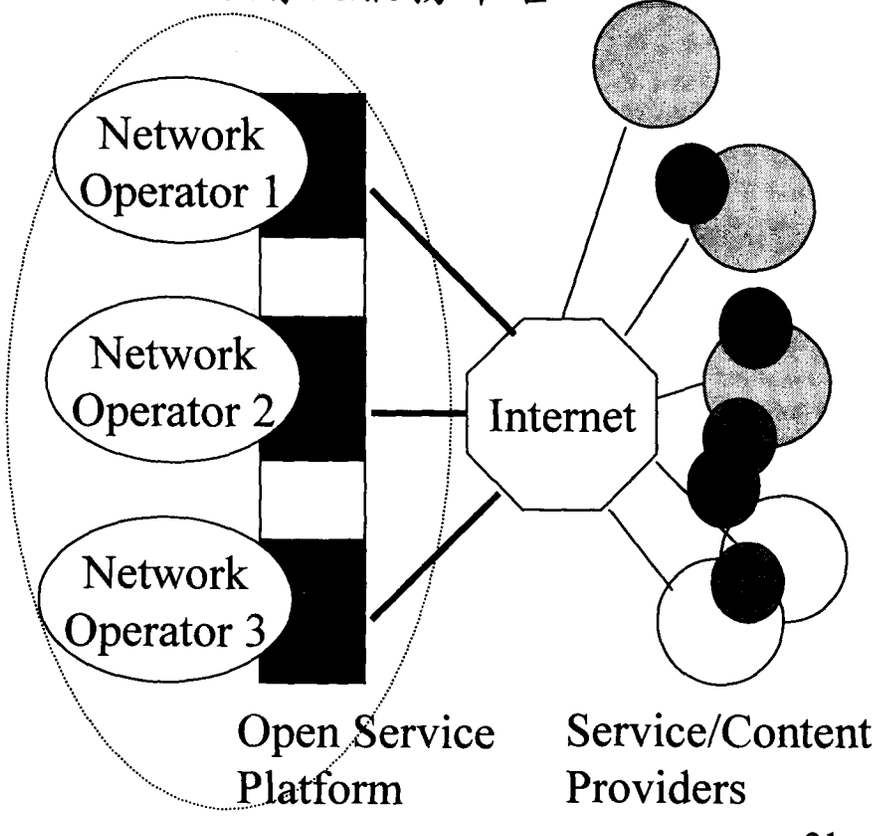
- Provide an extendible and scalable architecture that allows for inclusion of new service capability features and service capability servers.
- API is general; independent of
  - vender specific solution;
  - Programming language;
  - Operating systems.
- Object-oriented techniques, like CORBA, can be used to achieve above goals.

Why needs OSA?

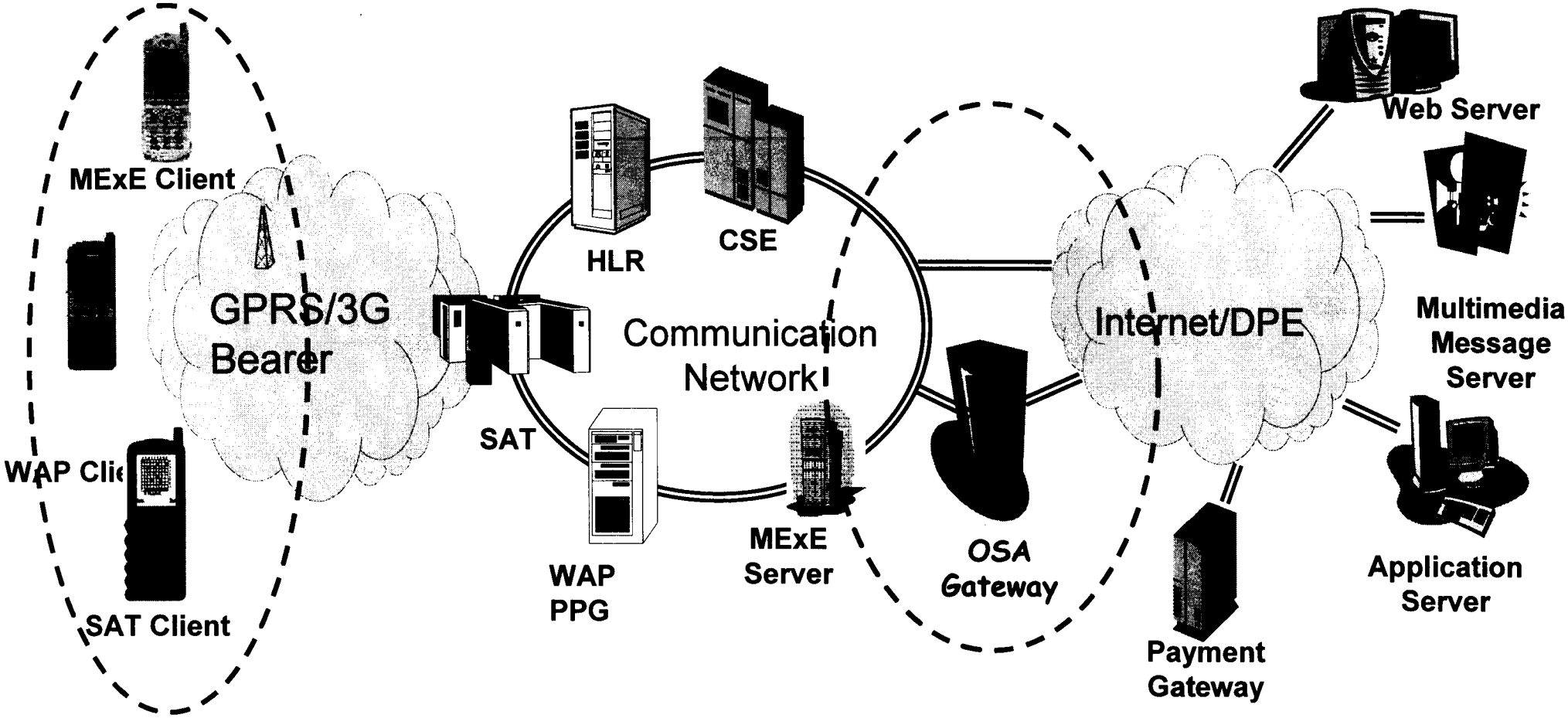
- Network Operator多，環境不一服務業者投入意願難提昇



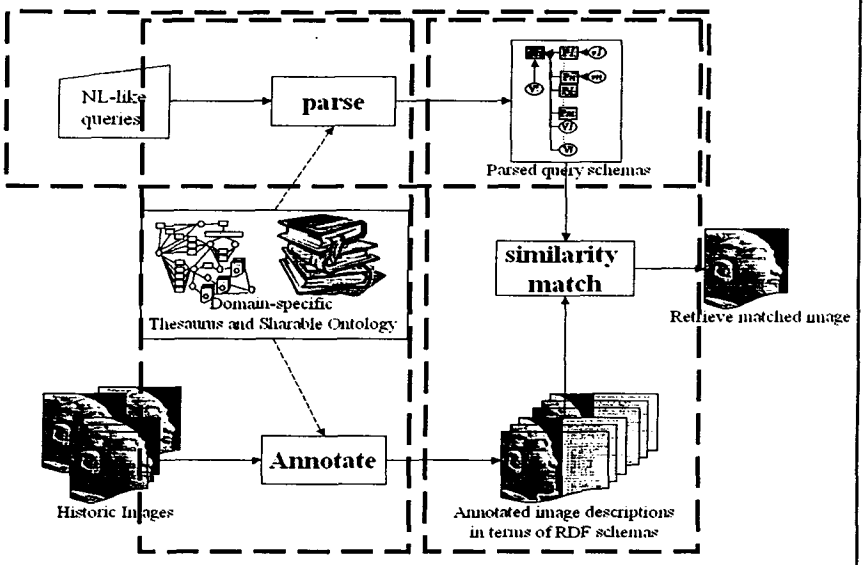
- 整合各Operator，提供相同之開放服務平台



OSA Capability



Processes of ontology-based image retrieval



Sharable Ontology & Thesaurus

- Ontology
  - Based on RDF Schema
  - Describe the Relations between classes
  - Currently implemented 6 classes and about 100 properties.
- Thesaurus
  - General term: about 70'000 terms in 13 categories.
  - Domain term: add about 300 terms in historical domain of Qin terracotta soldiers.

行動式資訊交換開放平台技術

Service Terminal



MExE Client

* **Embedded Browser**
* Mobile Information Device Profile
* J2ME Core APIs

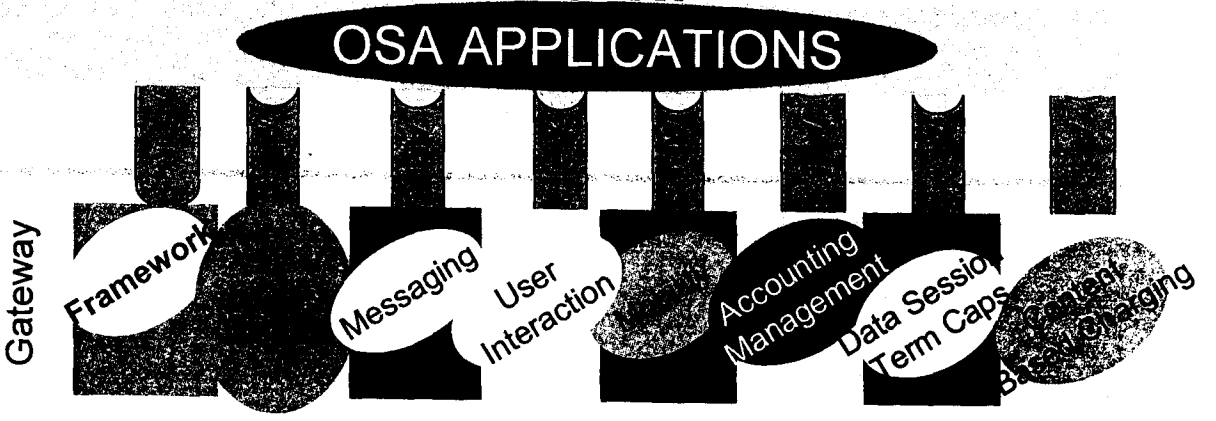
KVM

RTOS

Common Service Platform



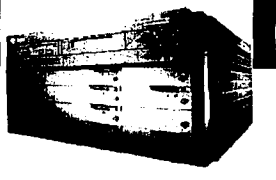
Open Service Access



Framework SCFs
- *Authentication
- *Authorization
- *Discovery
- *Registration
- *Notification
- *Integrity Management
- *Security/Privacy
- *Public Key Infrastructure (extension)

Network SCFs
- Call Control
- *Data Session Control
- User Location
- User Status
- *Terminal capabilities
- *User Profile Mngt.
- Billing

MExE Server



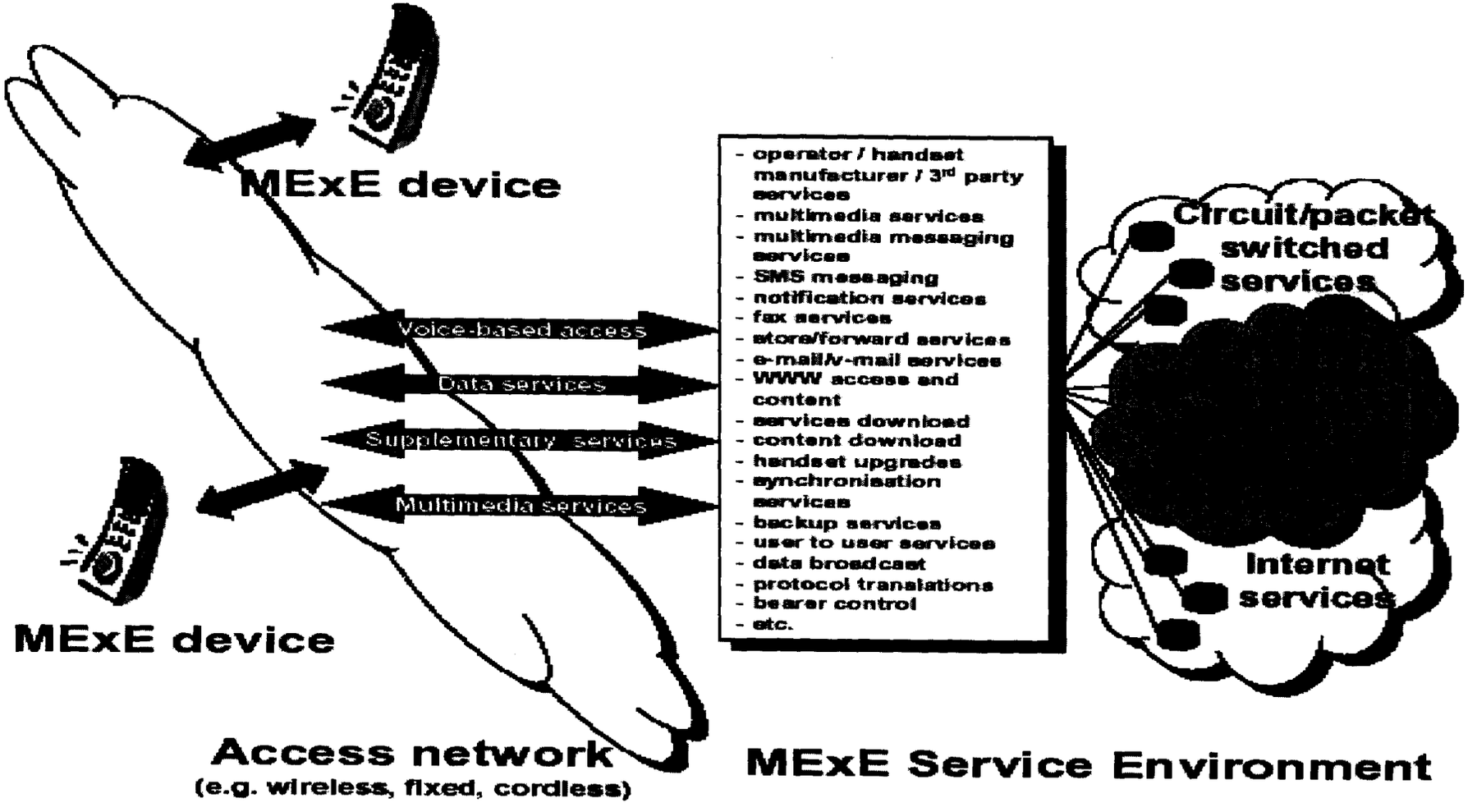
WAP PPG

MExE

- **Mobile Execution Environment**
  - details a flexible and secure application environment for 3G (and 2G+) mobile devices.
  - MExE consists of a variety of current technologies and incorporates both WAP and Java.
  - It also includes a framework which specifies:
    - Capability and content negotiation
    - User profiles
    - Management of services
    - Virtual home environment
    - User control of application connections
    - Journalling of network events

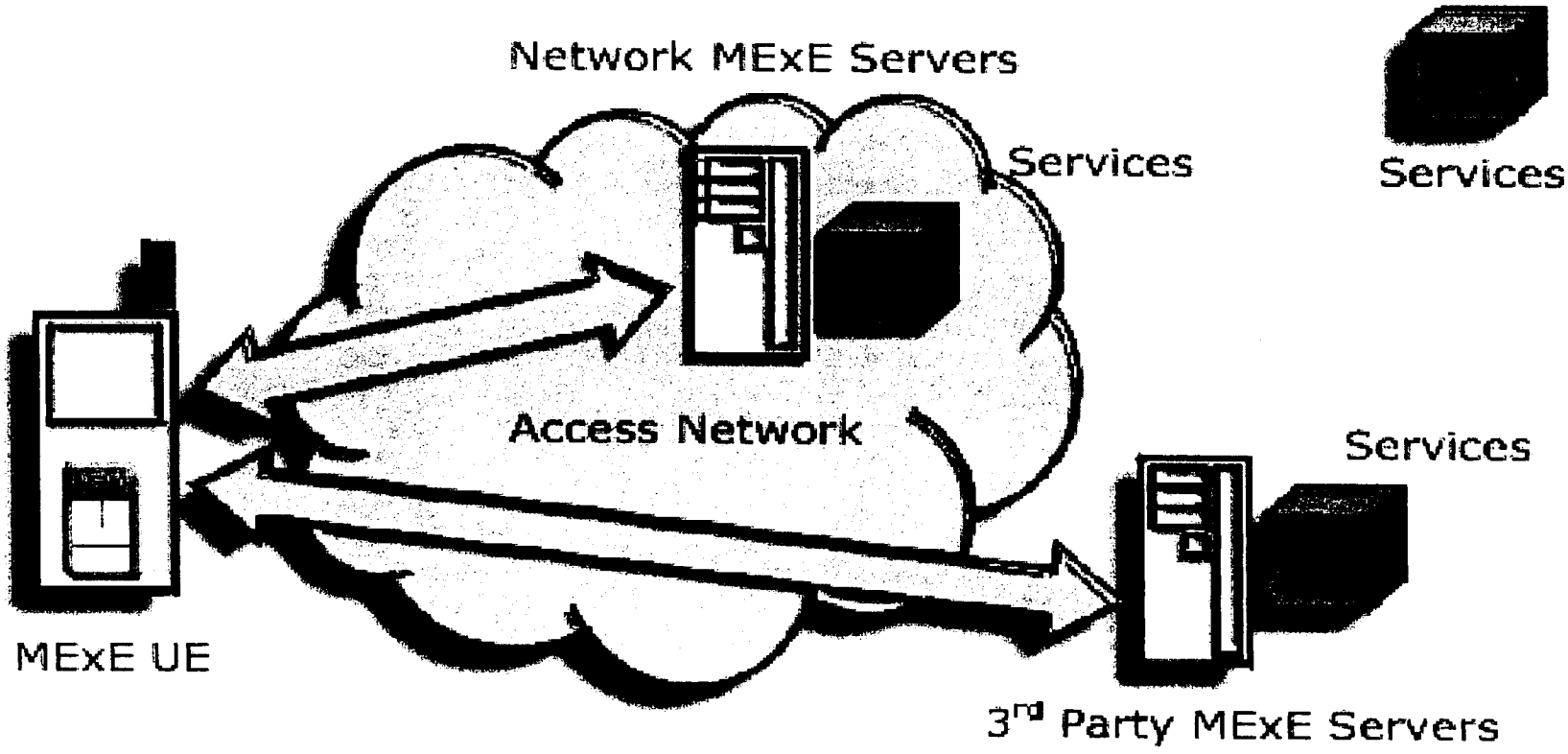
MExE Aspects

- High level architecture



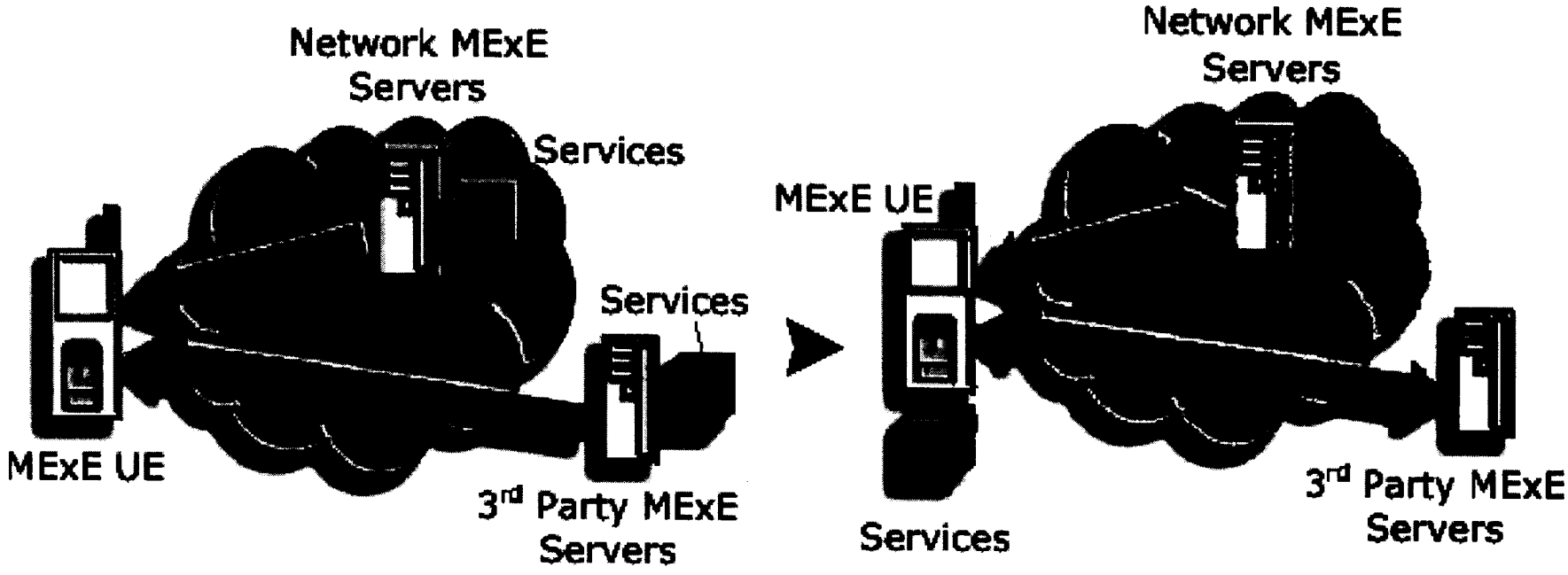
MExE Services (1)

- Access to MExE Services
  - –Services execute on a remote server



MExE Services (2)

– Service client downloaded to MExE MT

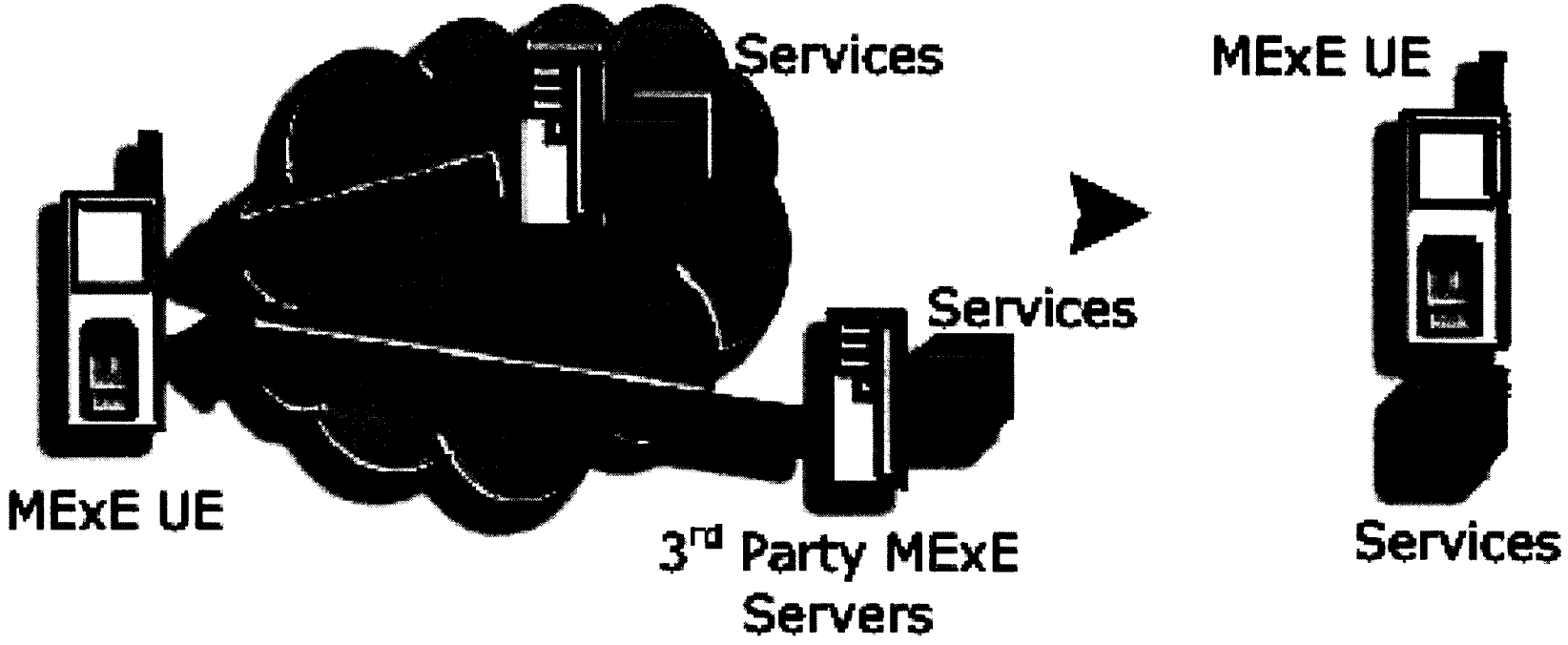


Step1:
MExE UE Downloads New Services

Step2:
New Services Interact with Servers

MExE Services (3)

–Standalone service downloaded to MExE MT

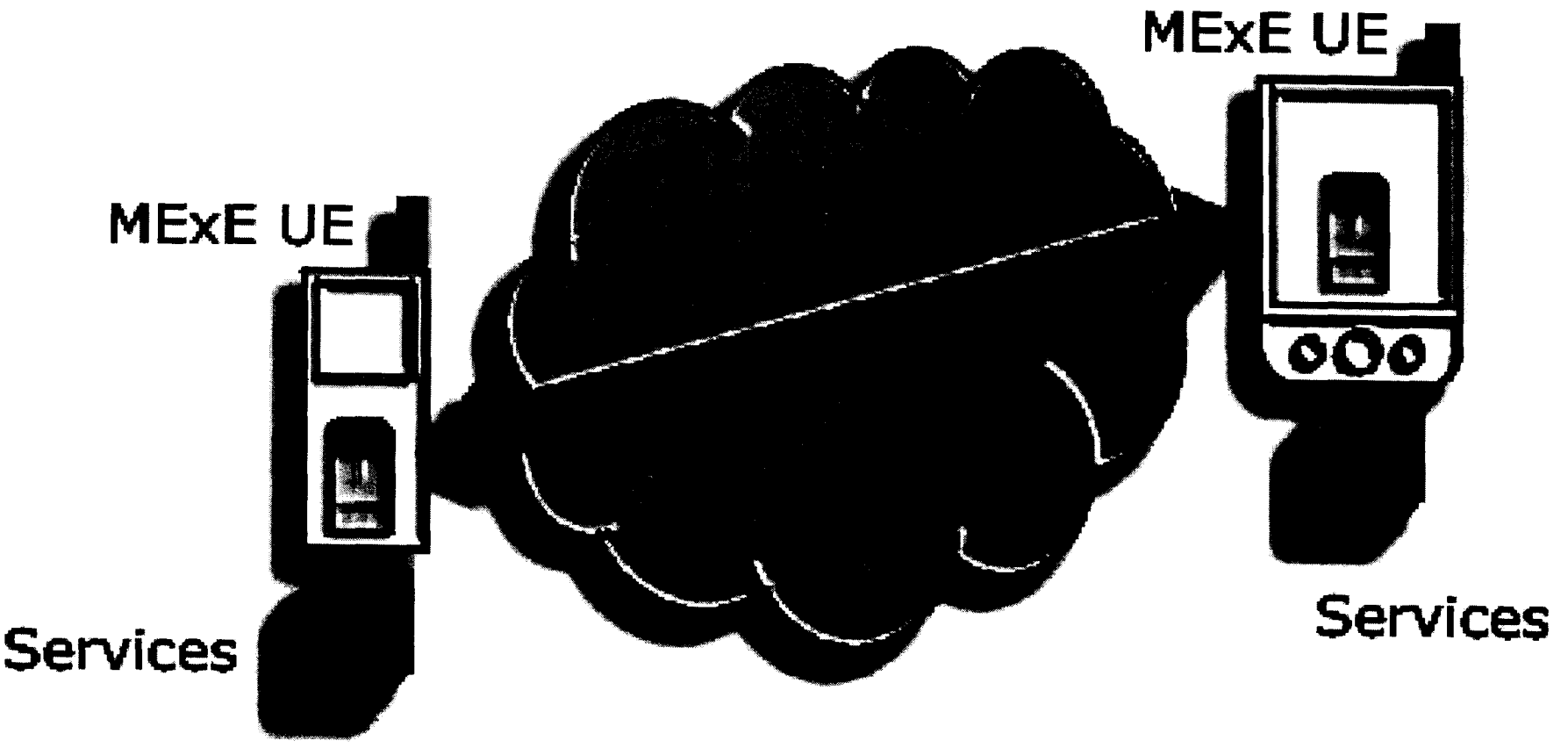


Step1:
MExE UE Downloads New Services

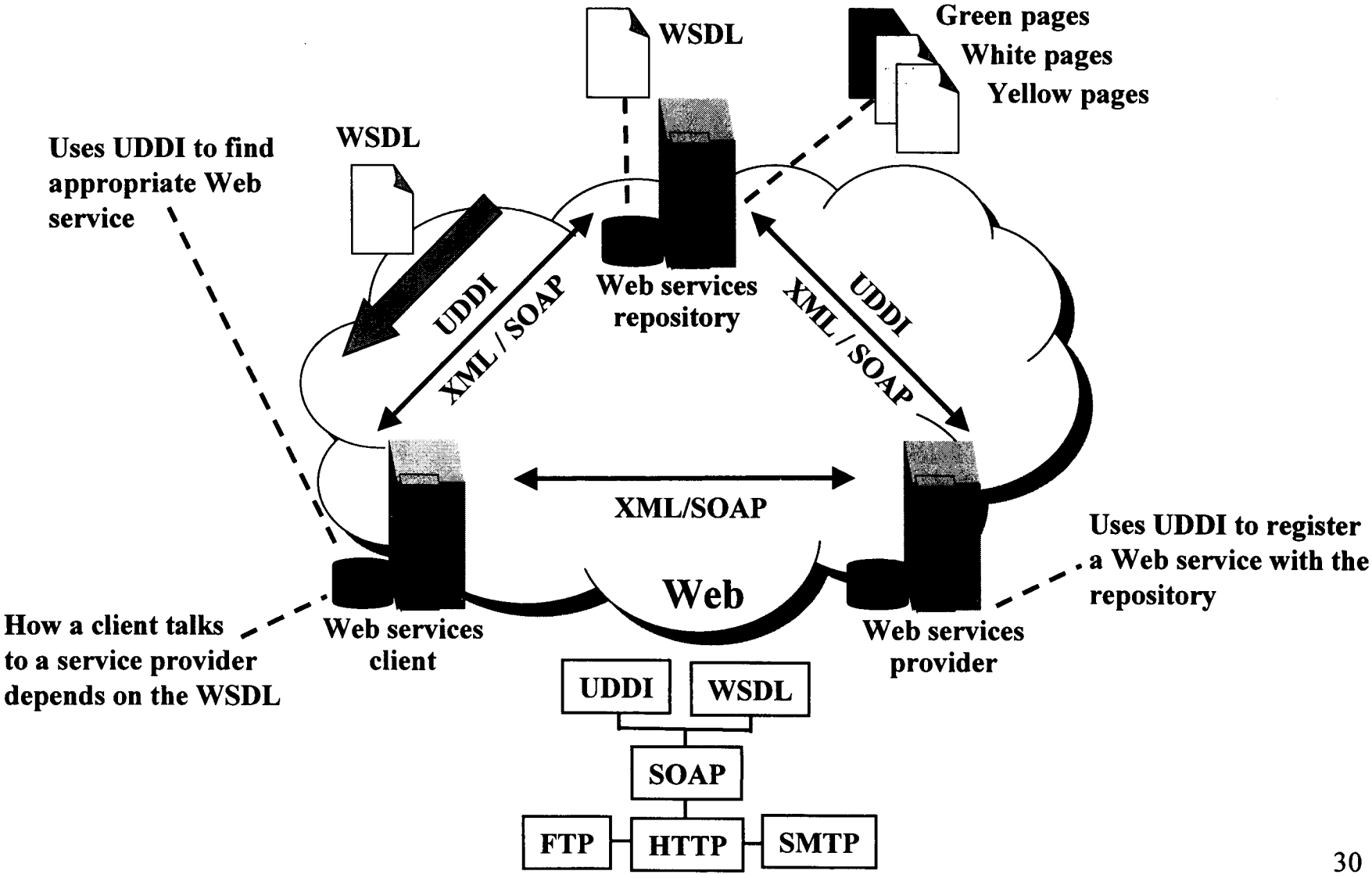
Step2:
Services Execute on the UE

MExE Services (4)

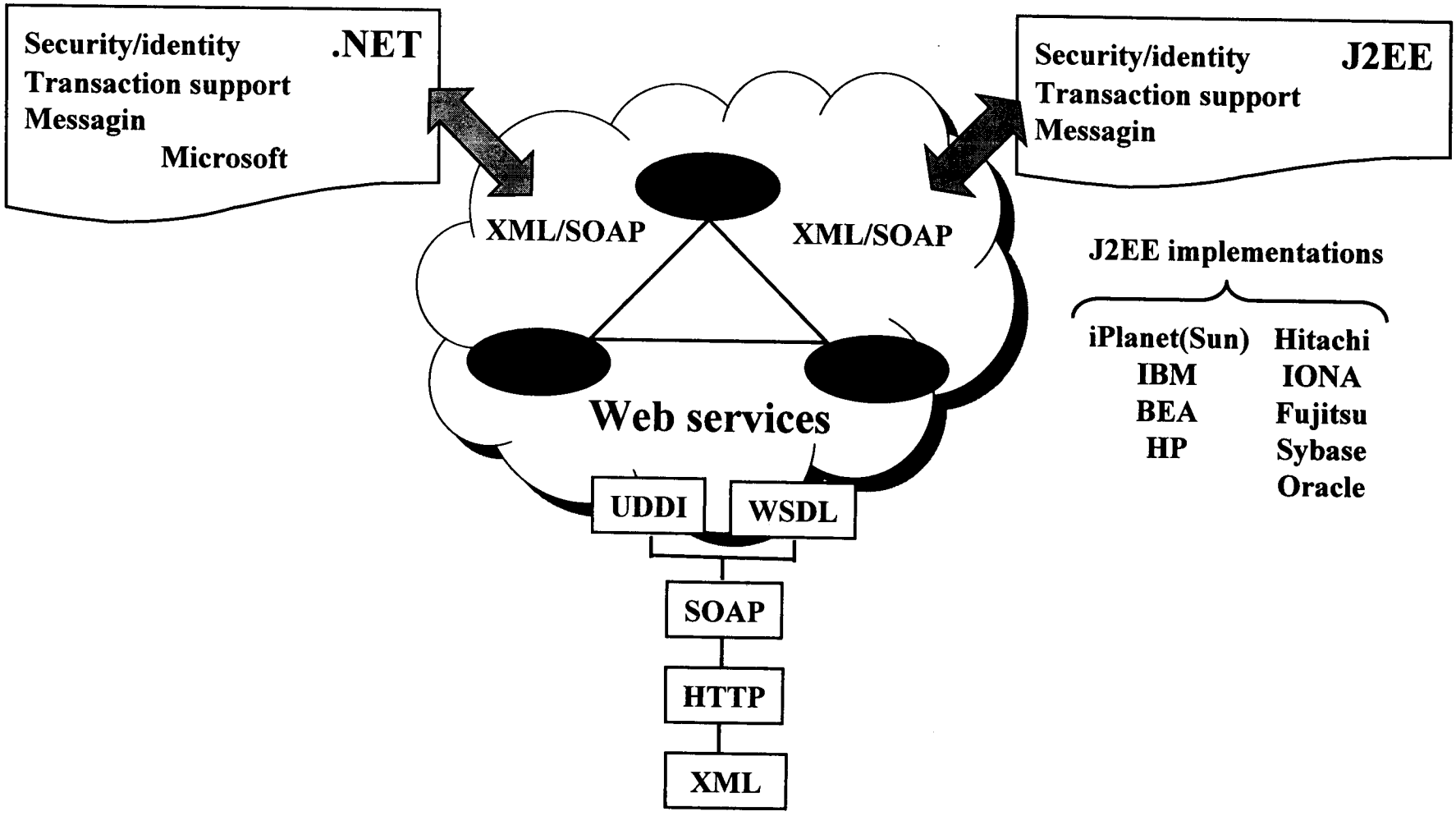
–MExE MT to MExE MT services



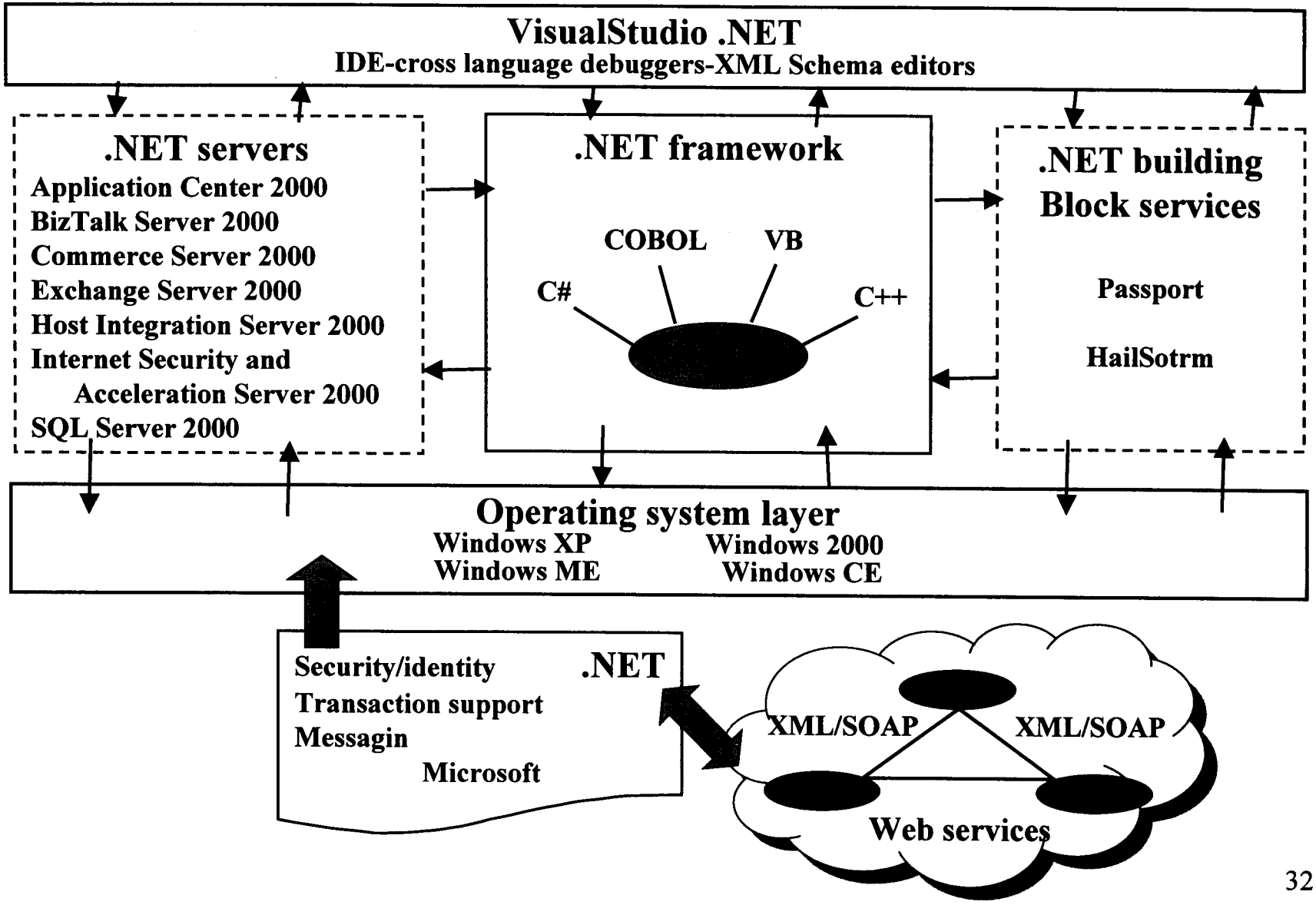
Web Services Architecture



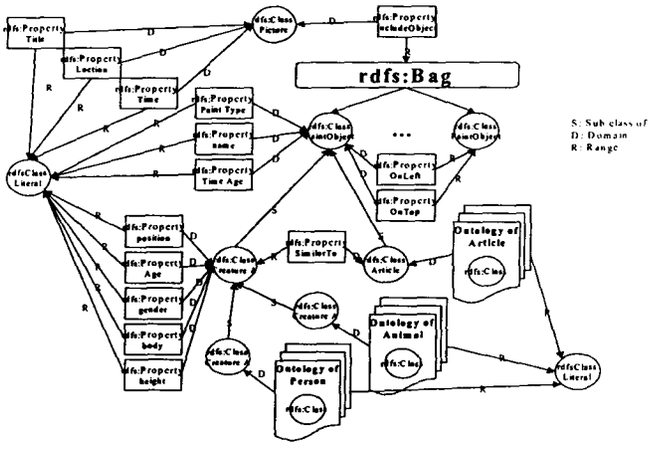
.NET and J2EE


Security/identity
Transaction support
Messagin
.NET
Microsoft
Security/identity
Transaction support
Messagin
J2EE
XML/SOAP
XML/SOAP
Web services
UDDI
WSDL
SOAP
HTTP
XML
J2EE implementations
iPlanet(Sun)
IBM
BEA
HP
Hitachi
IONA
Fujitsu
Sybase
Oracle

Microsoft .NET Platform



Sharable domain ontology for terracotta warriors, horses and related articles (in Graphic representation)

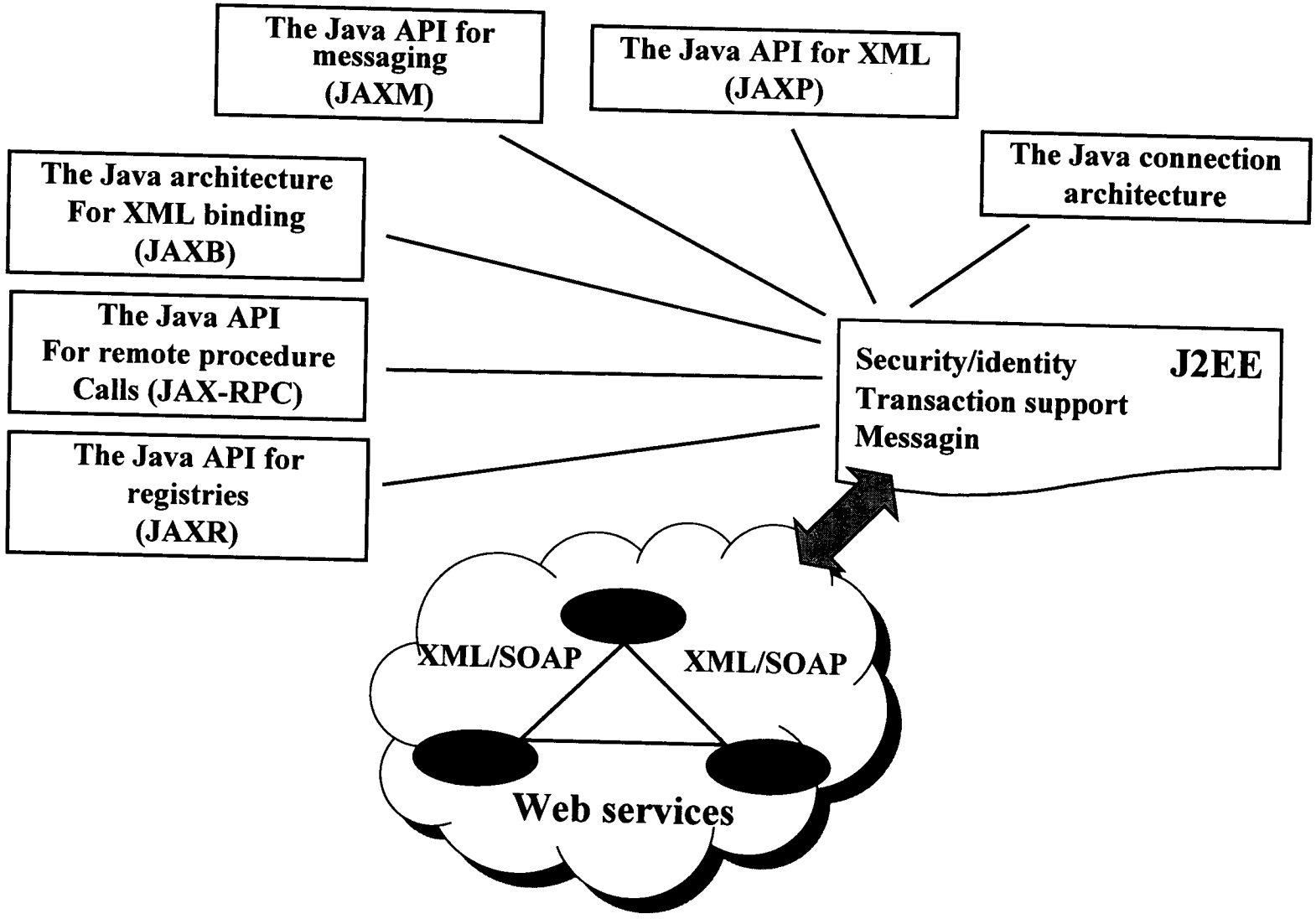


An instance of the sharable domain ontology (in RDFS)

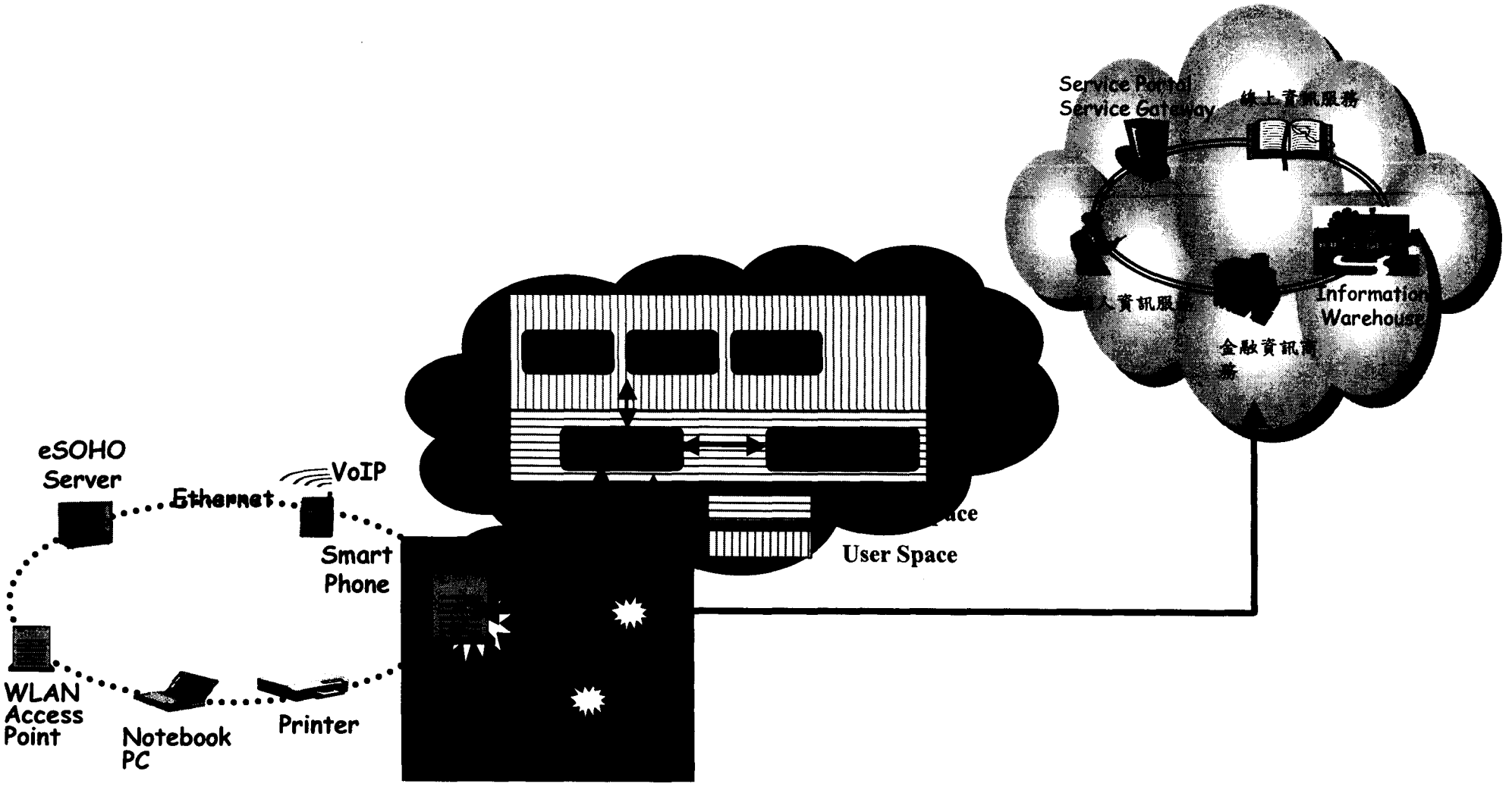
```
<rdfs:Class rdf:ID="生物">
  <rdfs:subClassOf rdf:resource="#圖形物件"/>
</rdfs:Class>
<rdf:Property ID="姿勢">
  <rdfs:domain rdf:resource="#生物"/>
  <rdfs:range rdf:resource="???:姿勢集合"/>
</rdf:Property>
<rdf:Property ID="年齡">
  <rdfs:domain rdf:resource="#生物"/>
  <rdfs:range rdf:resource="???:數字集合"/>
</rdf:Property>
<rdf:Property ID="性別">
  <rdfs:domain rdf:resource="#生物"/>
  <rdfs:range rdf:resource="???:性別集合"/>
</rdf:Property>
<rdf:Property ID="體型">
  <rdfs:domain rdf:resource="#生物"/>
  <rdfs:range rdf:resource="???:體型集合"/>
</rdf:Property>
<rdf:Property ID="身高">
  <rdfs:domain rdf:resource="#生物"/>
  <rdfs:range rdf:resource="???:數字集合"/>
</rdf:Property>

<rdfs:Class rdf:ID="物品">
  <rdfs:subClassOf rdf:resource="#圖形物件"/>
</rdfs:Class>
<rdf:Property ID="模擬意">
  <rdfs:domain rdf:resource="#物品"/>
  <rdfs:range rdf:resource="#生物"/>
</rdf:Property>
<rdf:Property ID="敘路">
```

J2EE Web Services pack



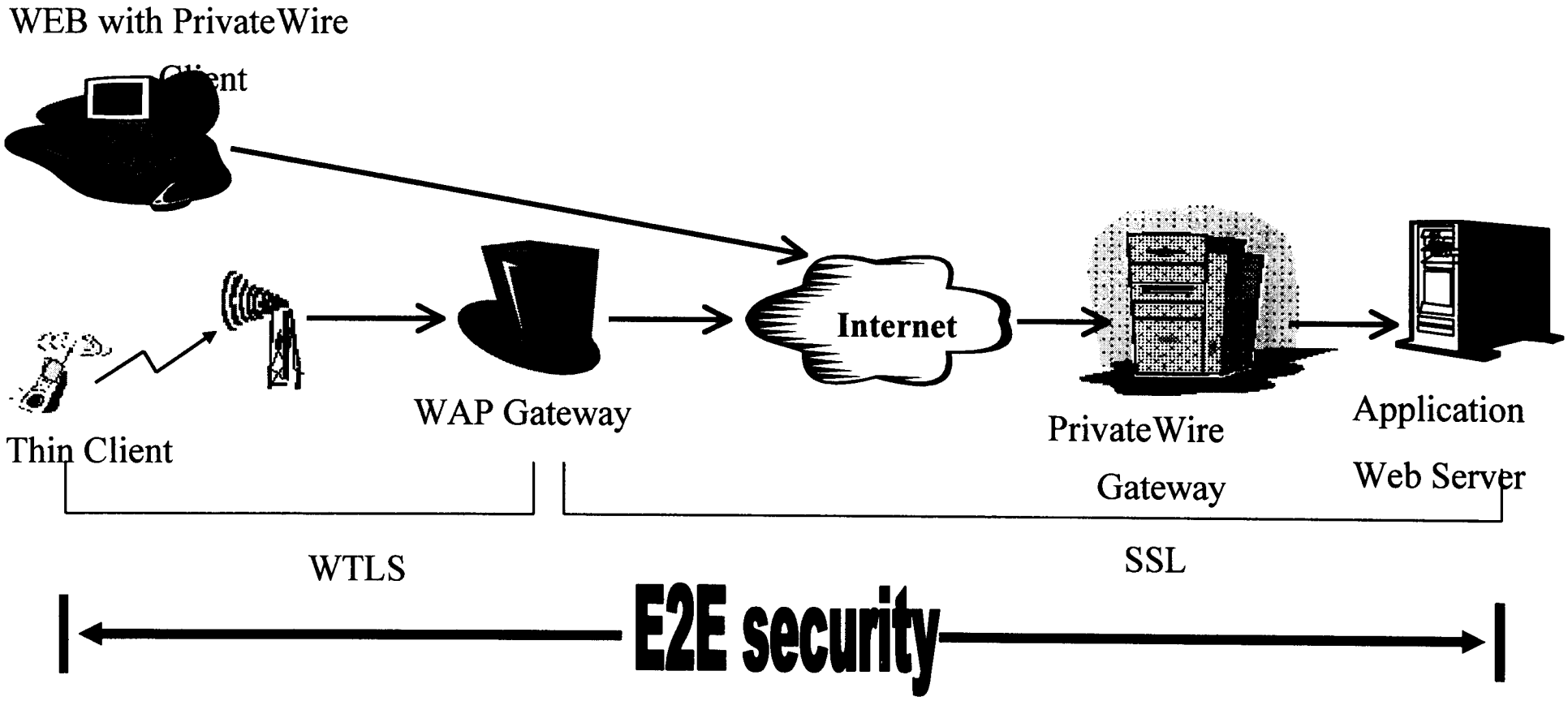
P2P/Web Srvice Gateway



P2P/Web Srvice Gateway(續)

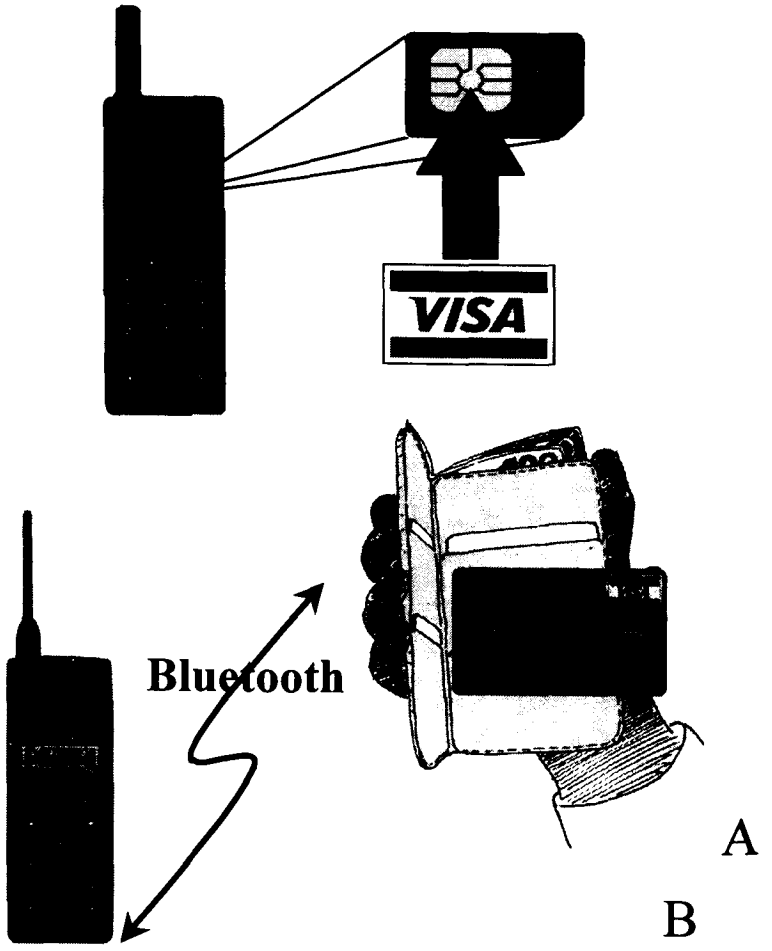
- Proxy & QOS
- Personalization
- Content Filtering
- Service Mapping
  - Export mapping
  - Import Services
- Transformation
- Service Publication & Lookup
- Security & Management
- Service Provision/Brokering
- NAT/Firewall

M-Commerce Security WTLS vs SSL
- WAP Solution



Wireless Identity Module(WIM) Card
(Smart Card + SIM Card)

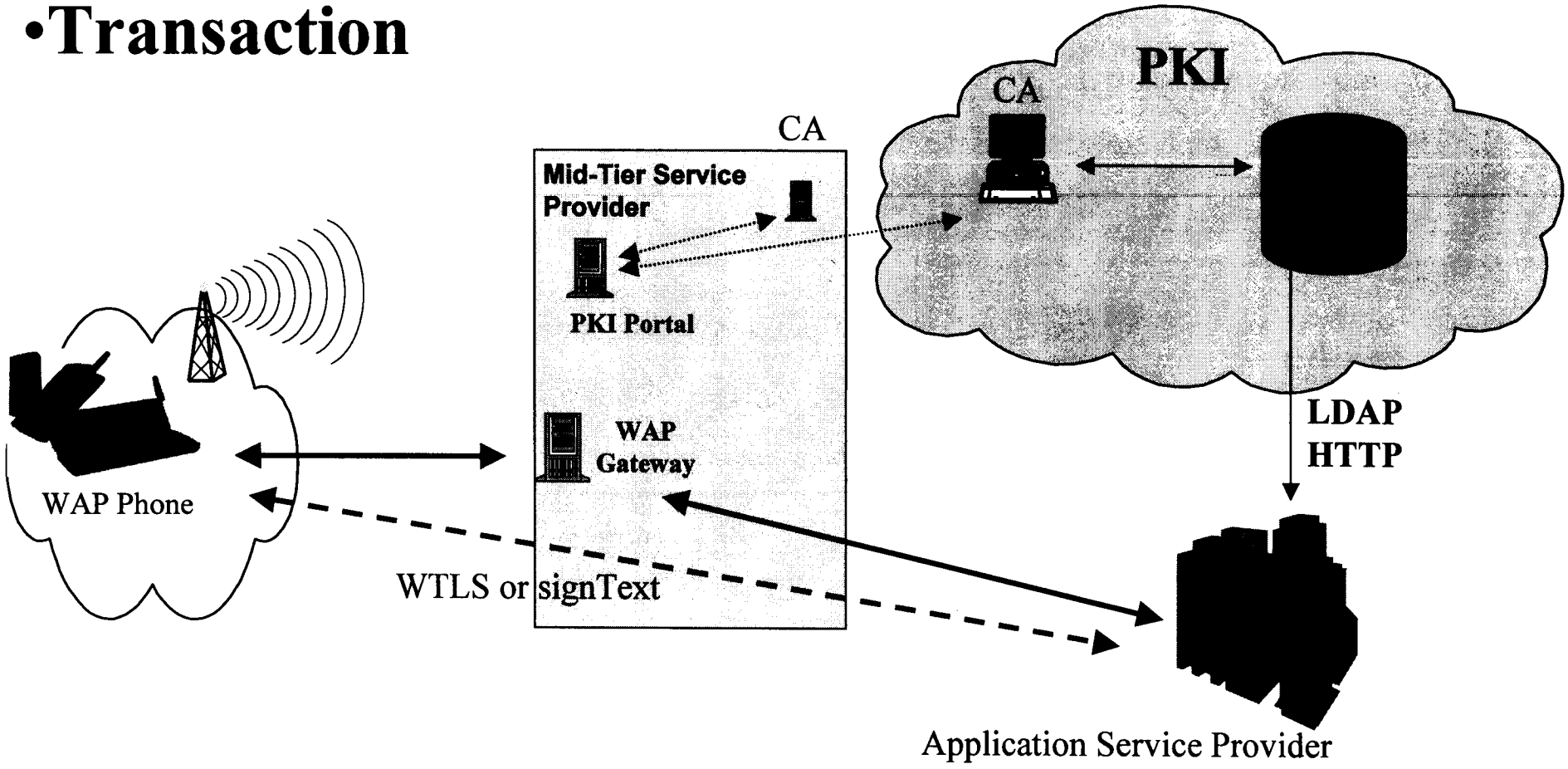
Application(EC, Push Application)				
WAP(1.2,1.3,2.0) Application			Non-WAP Application	
WPKI (draft)	WTLS	WML	SSL3.0(TLS)	S/MIME 3.0
PKCS#15 (Information format)	Unwrapping key,Digital Signature,hash Verify Certificate,Access Control…			
ISO 7816				



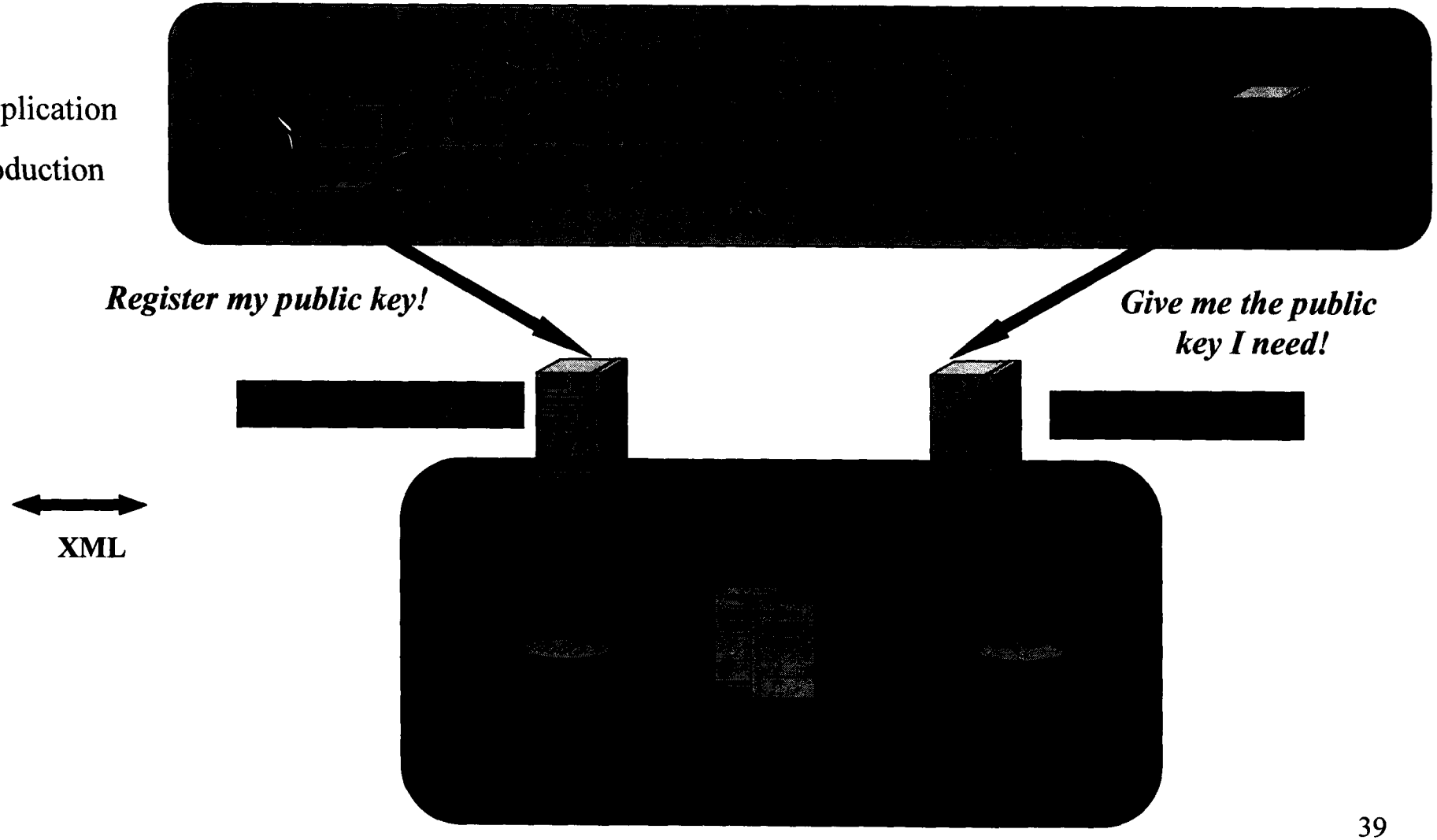
☑ Privacy ☑ Integrity ☑ Authentication ☑ Non-Repudiation

WPKI Architecture

•Transaction

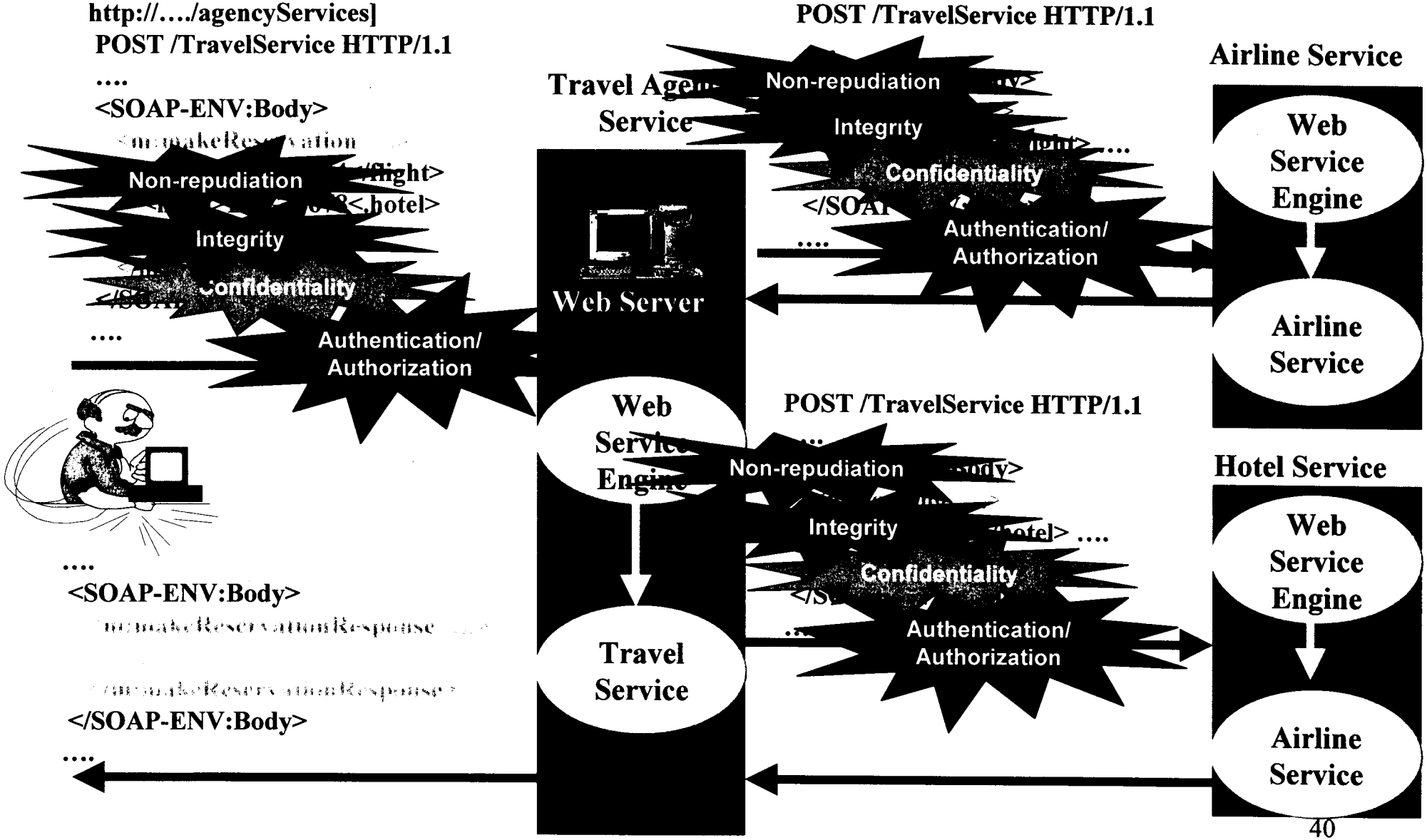


XML Key Management System(XKMS)

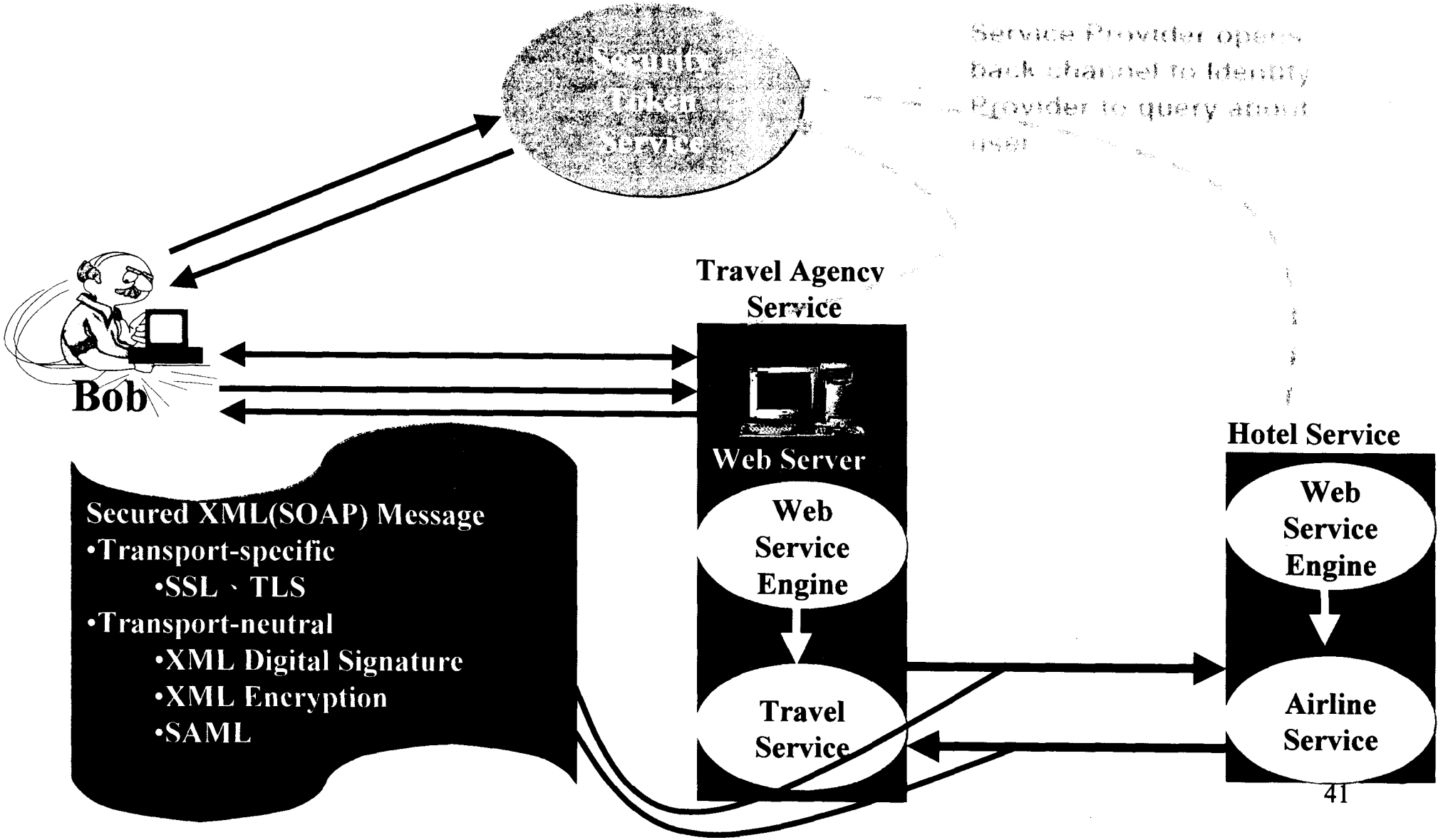


工業技術研究院
電腦與通訊工業研究所
Industrial Technology Research Institute
Computer & Communications Research Laboratories

Web Service Security Issues



Web Service Security: Possible Solution



Web Service Security in The World

- TrustBridge
  - 係根據微軟與IBM、VeriSign聯手發展的網路服務安全技術，著重於一種稱為 「網路服務安全」(WS-Security)的規格。
- Liberty Alliance Project
  - 自由聯盟計畫（www.projectliberty.org）是一個嶄新的組織，成立目的為了建構開放式的聯合單一登入Single sign-on）解決方案，符合數位經濟以任何設備上網的需要。

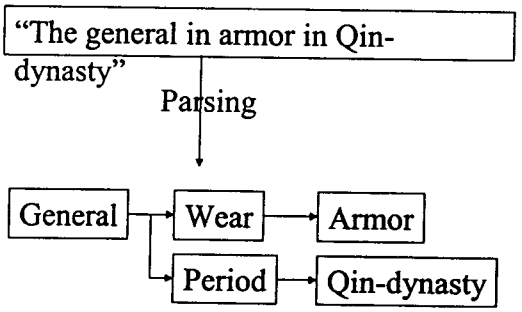
An annotated image of a side view of a Qin terracotta warrior's head




```
<?xml version='1.0' encoding='big5'?>
<!DOCTYPE rdf:RDF [<!ENTITY p 'http://140.114
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-
<rdf:RDF xmlns:p="&p;"xmlns:rdf="&rdf;">
<p:圖片 rdf:ID="圖片48798">
  <p:Title>戰士的頭部</p:Title>
  <p:時間>1983</p:時間>
  <p:地點>一號坑 臨潼</p:地點>
  <p:IncludeObject rdf:resource="#bag"/>
</p:圖片>
<p:Article rdf:ID="Article1">
  <p:PaintType>照片</p:PaintType>
  <p:名稱>兵俑</p:名稱>
  <p:時代>秦朝</p:時代>
  <p:模擬為 rdf:resource="#person1"/>
  <p:用途>陪葬</p:用途>
  <p:材質>陶</p:材質>
  <p:數目>1</p:數目>
  <p:顏色>灰色</p:顏色>
  <p:形狀>人形</p:形狀>
</p:Article>
<p:person rdf:ID="person1">
  <p:性別>男性</p:性別>
  <p:國籍>中國秦朝</p:國籍>
  <p:身分>士兵</p:身分>
  <p:職業>軍人</p:職業>
  <p:頭 方向>右</p:頭 方向>
  <p:頭 臉部 表情>見牙</p:頭 臉部 表情>
  <p:頭 臉部 鬍鬚>八字鬍</p:頭 臉部 鬍鬚>
  <p:頭 臉部 耳朵>突出</p:頭 臉部 耳朵>
  <p:頭 臉部 鼻子>小鼻子</p:頭 臉部 鼻子>
</p:person>
<rdf:Bag rdf:ID="bag">
  <rdf:li rdf:resource="#Article1"/>
</rdf:Description>
</rdf:RDF>
```

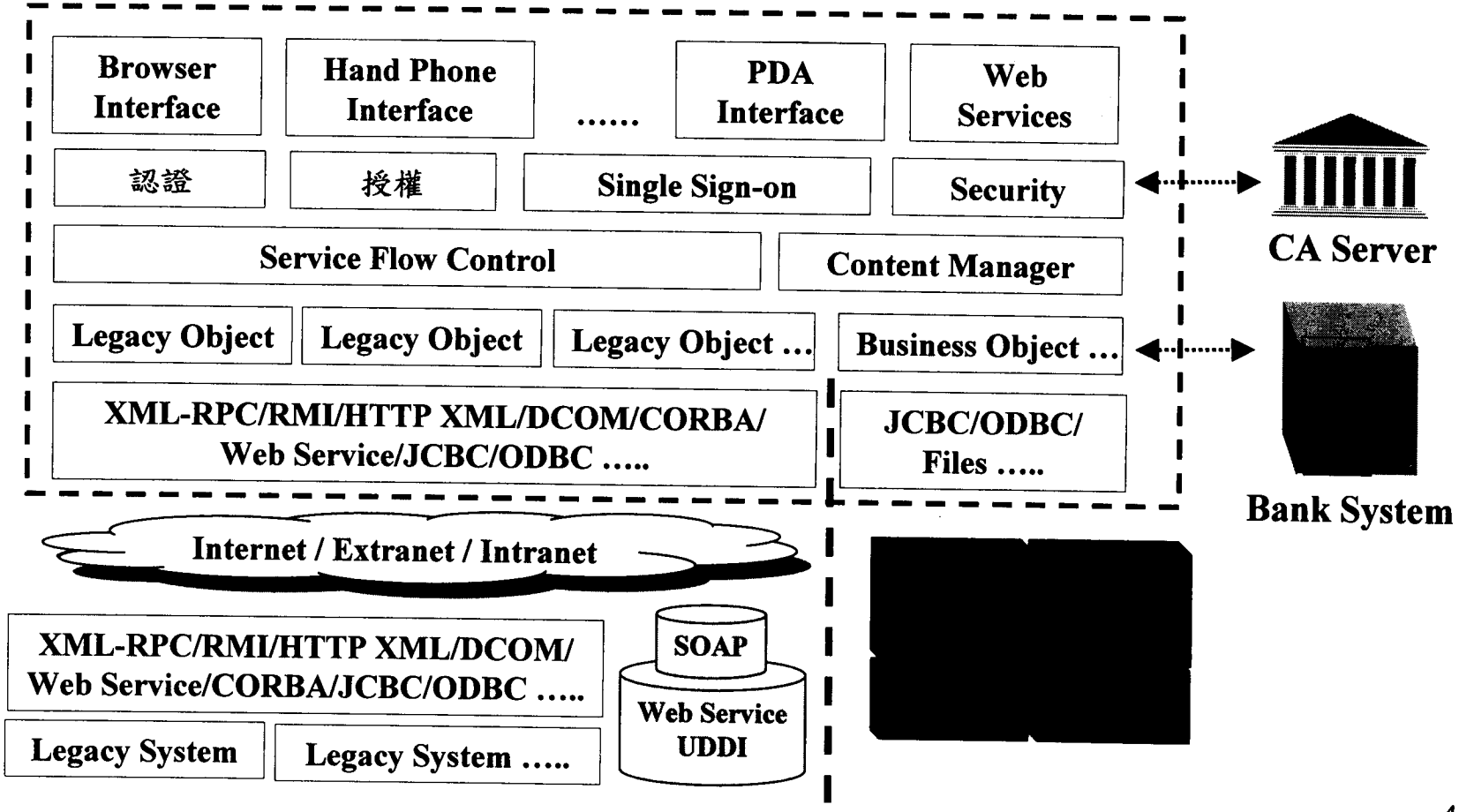
NL Query paring

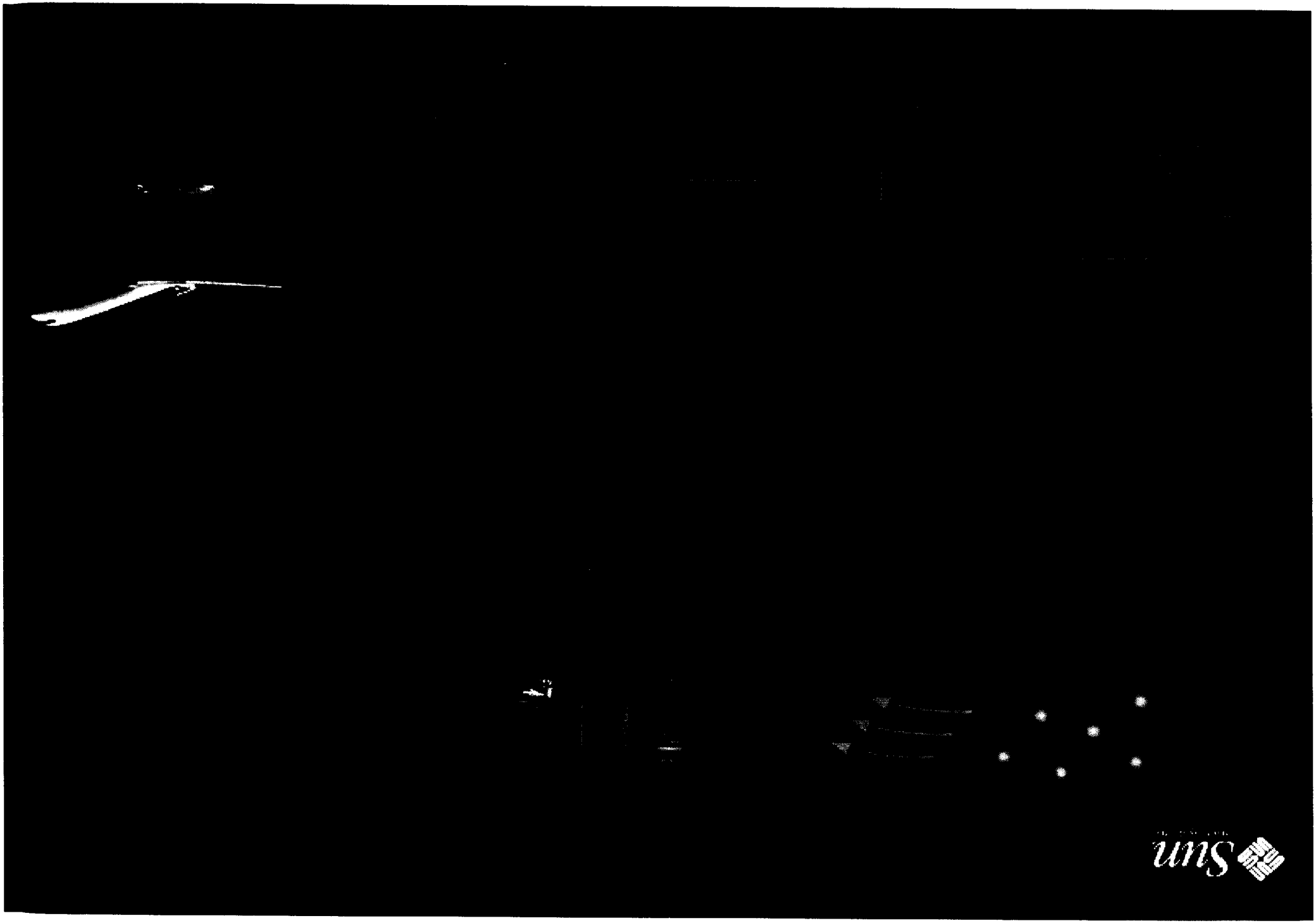
- ④ Users give the query in terms of a natural language phrase.
- ④ The system parses the query into the RDF format with the aid of ontology and thesaurus.



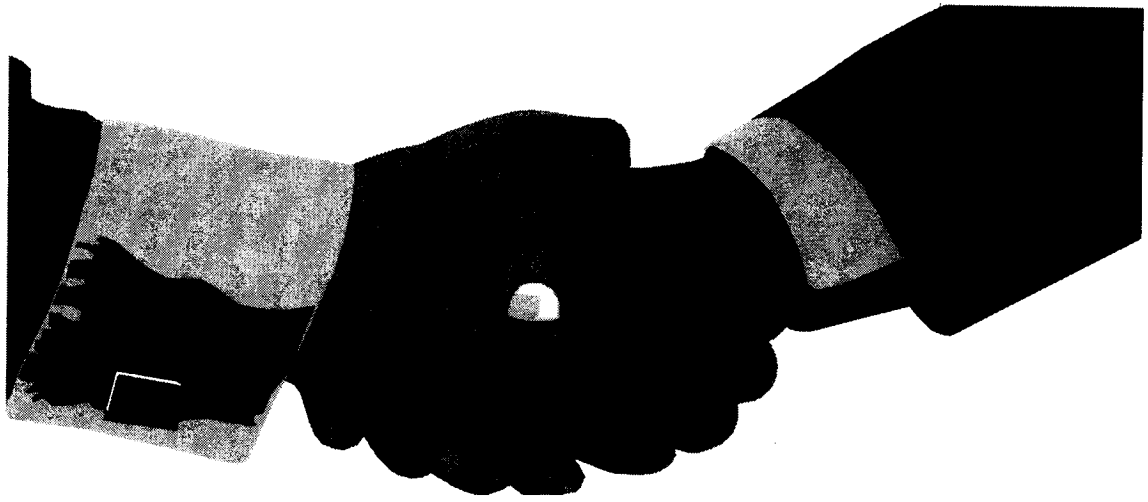
Web Service 整合服務架構







Thank you!



Mobile Web Services Workshop,NCKU, July 28, 2002

為行動服務而設計之共通行動終端平台

資策會
王台中

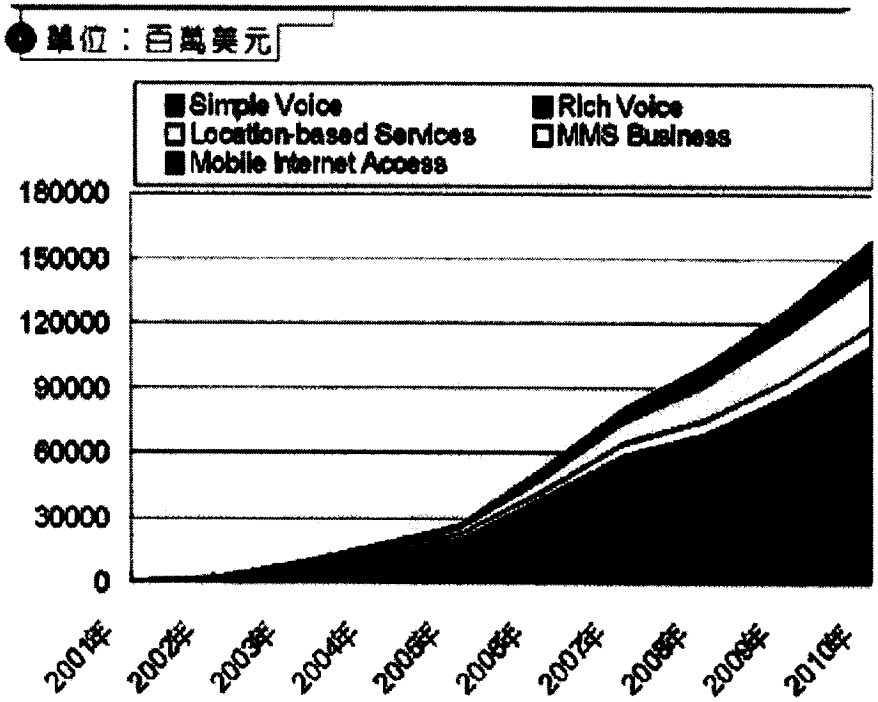
July 28, 2002

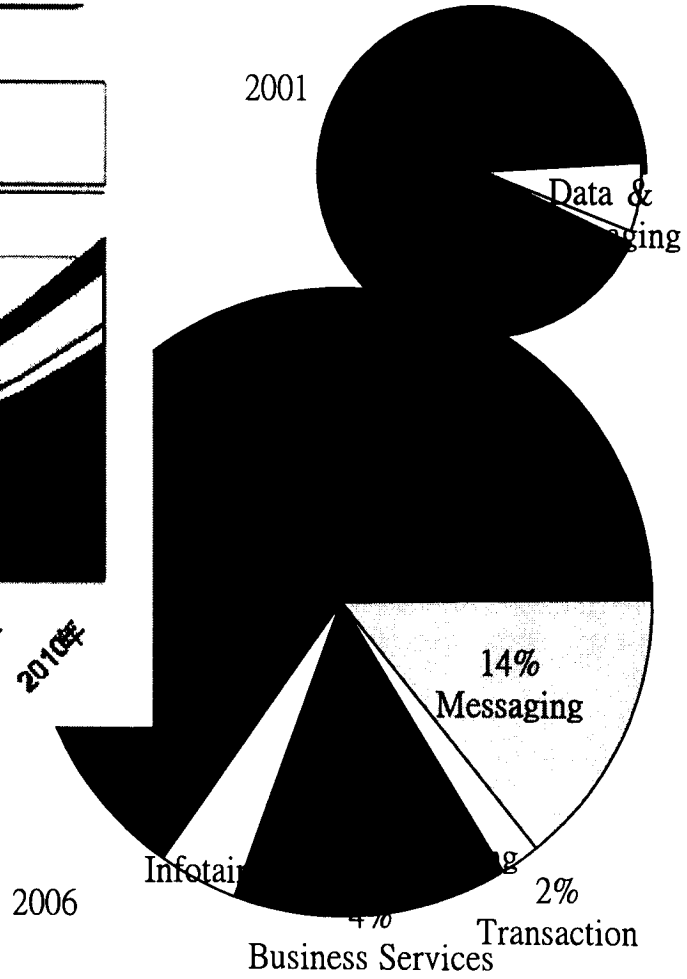
Contents

- 1.Mobile web service market opportunities
- 2.Mobile terminal system requirements
- 3.Mobile terminal platforms
- 4.Joint development / research for smart terminals

1. Mobile web service market opportunities

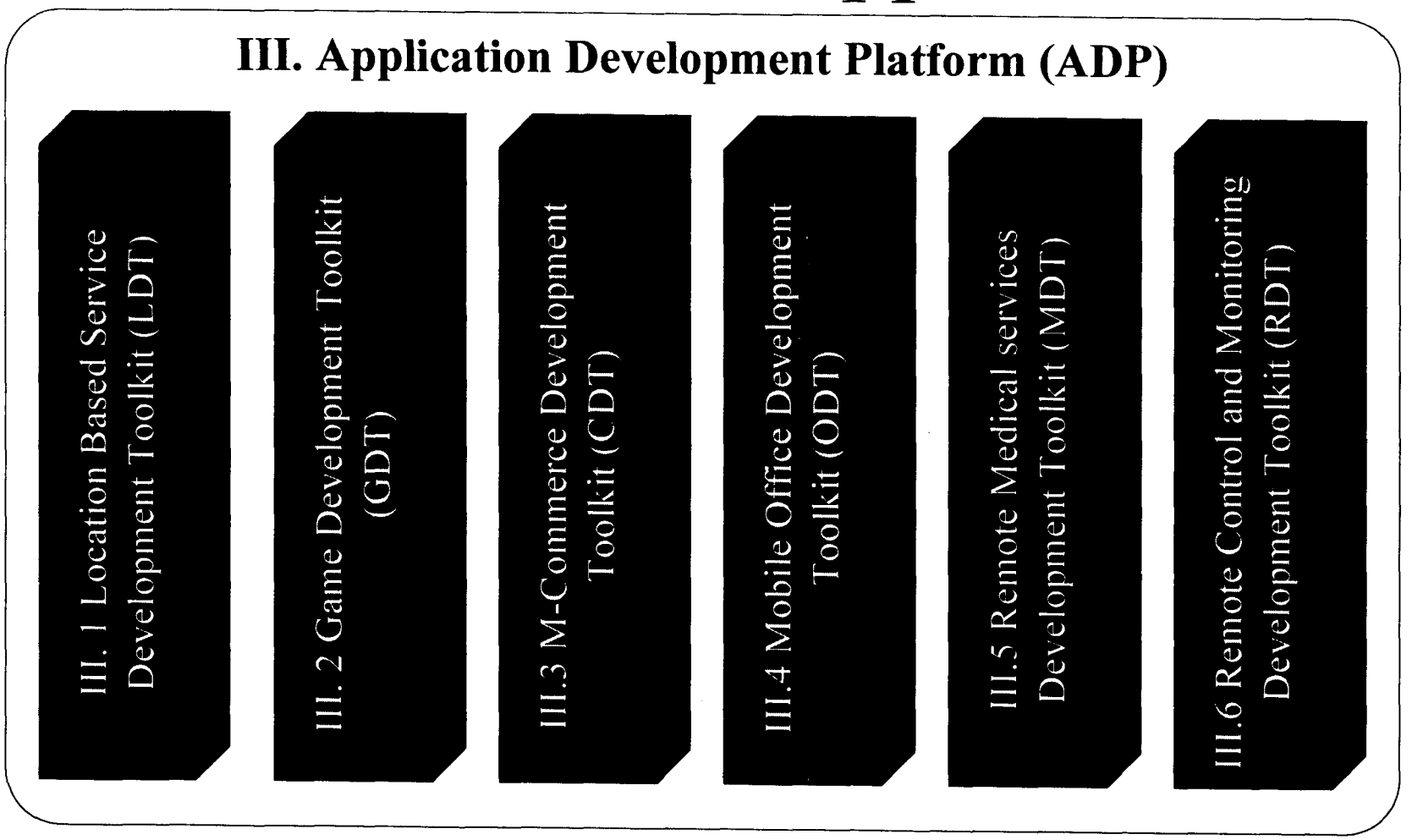
Mobile Service Market Share



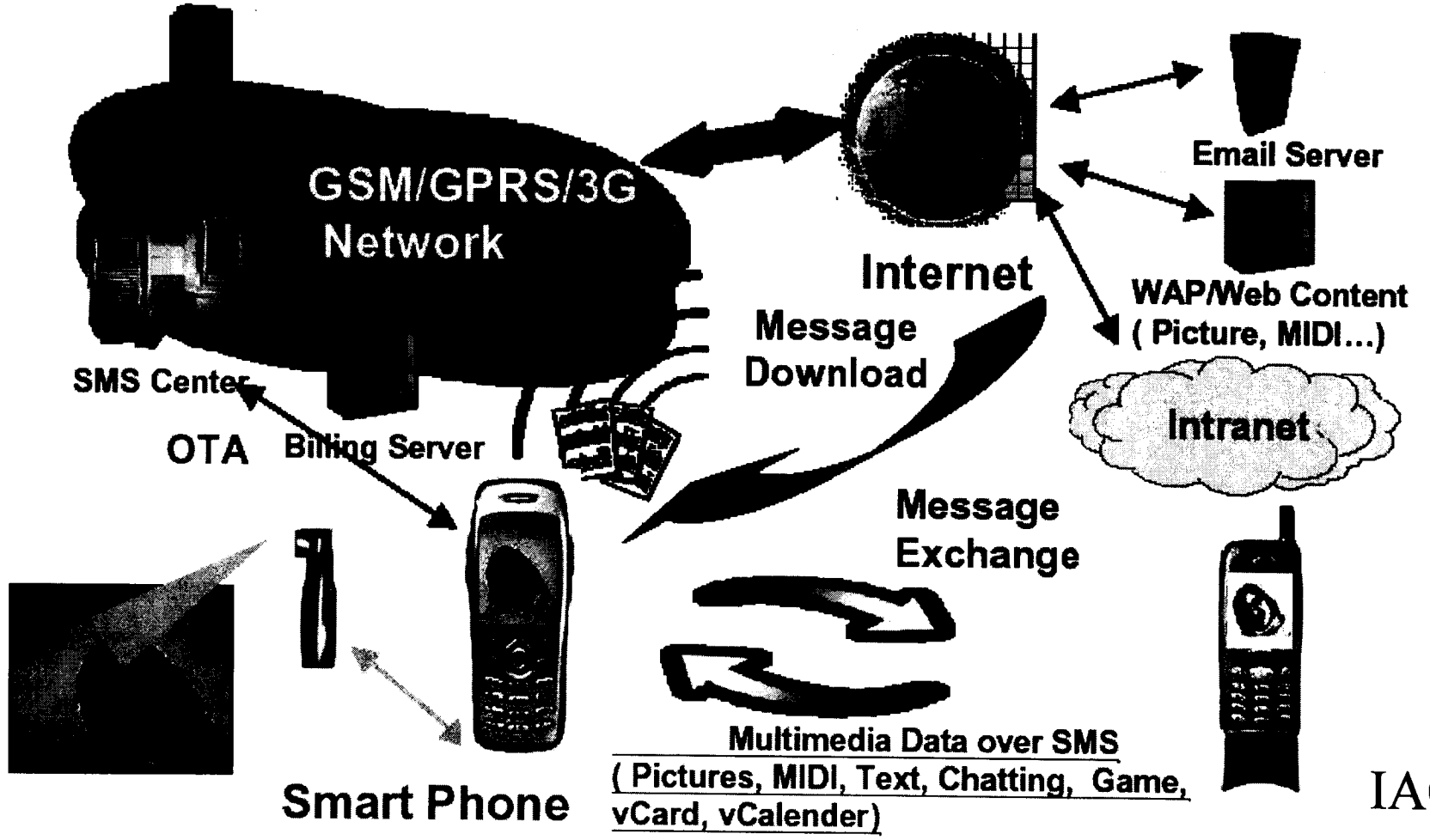


Data Source: Telecompetition, UMTS

Mobile Internet Applications



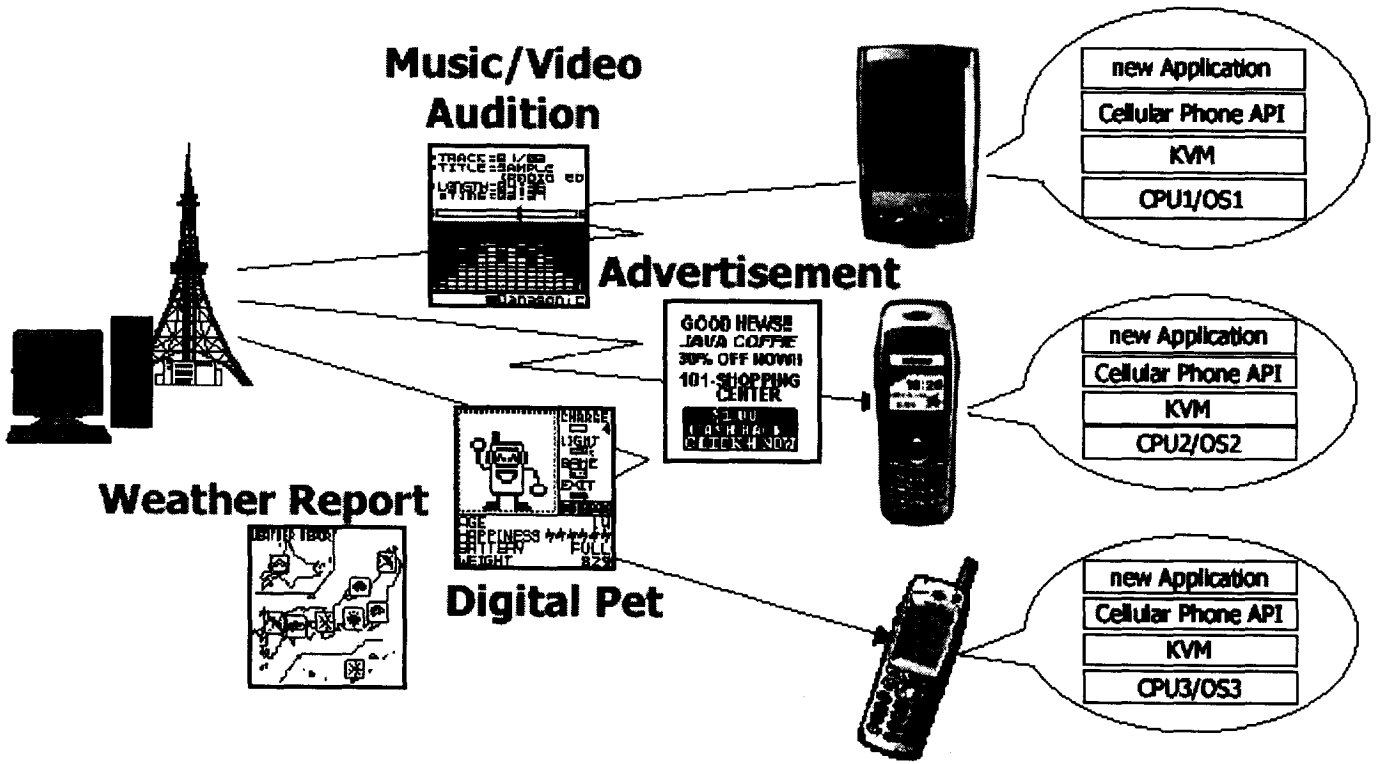
Messaging：MMS



IAC

Mobile Game

New Applications that Realize New Services are Downloaded and Run on "Profile"



IAC

NL Query paring (Naïve parsing Algorithm)

"秦代穿著盔甲的將軍" (The general in armor in Qin-dynasty)

Word segmentation

秦代 穿著 盔甲 將軍" (Qin-dynasty,Wear,Armor,General)

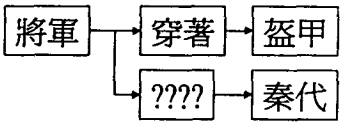
Property assignment

秦代 穿著 盔甲 將軍" (Qin-dynasty,Wear,Armor,General)

NL Query paring (Naïve parsing Algorithm)

秦代 穿著 盔甲 將軍"

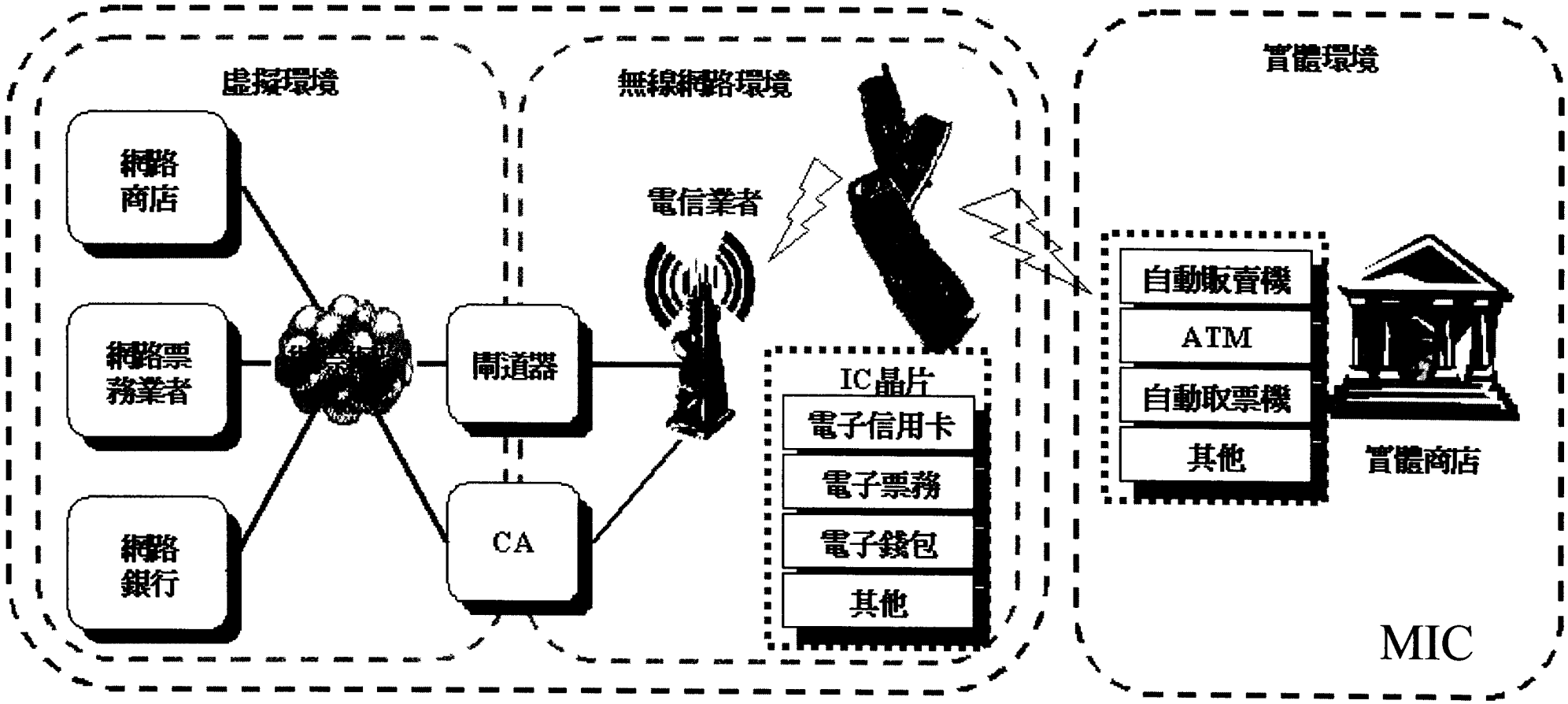
Backward matching



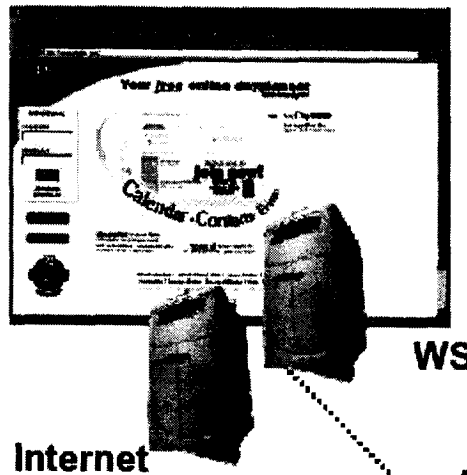
④ Disadvantage

– Too simple and easy to mismatch.

Mobile Commerce



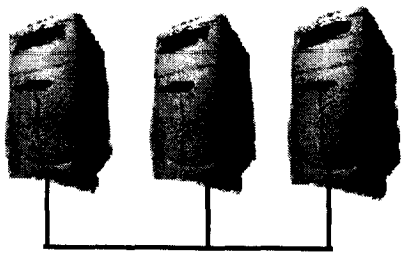
Mobile Office



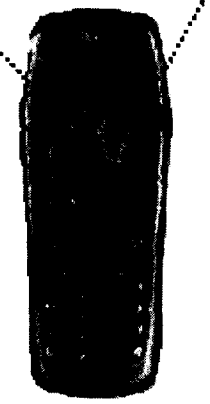
WSP

Internet
- Internet address book
- Internet calendar
- Internet mail
- Stock prices, etc

HTTP



Company LAN
- Company calendar
- Company Database
- Mail server
- Company files
- Own files (docs, backups)



Smart Phone

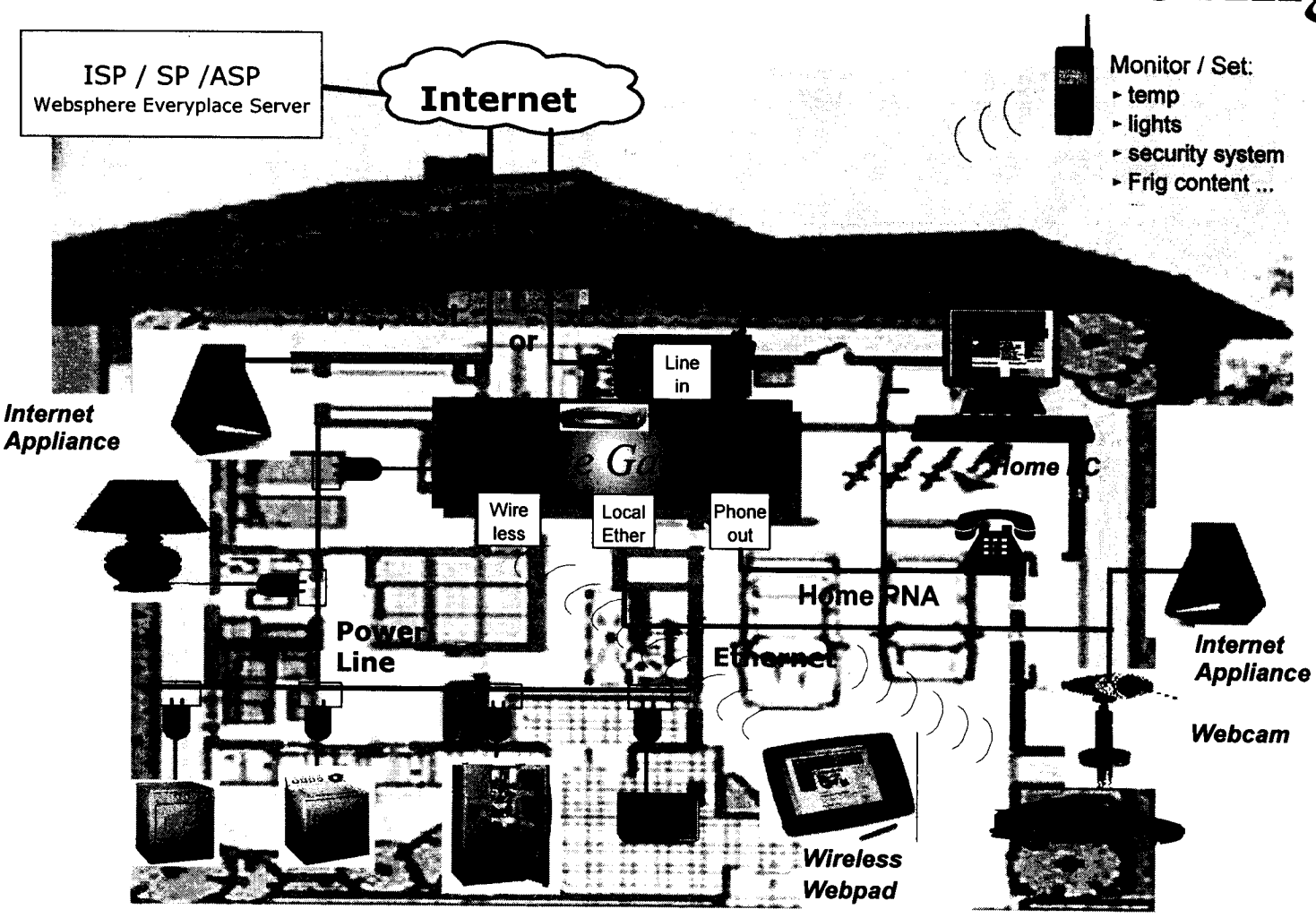
Bluetooth,
Infrared



PC
- Calendar
- Contacts
- Docs, Files

IAC

Remote Control and Monitoring



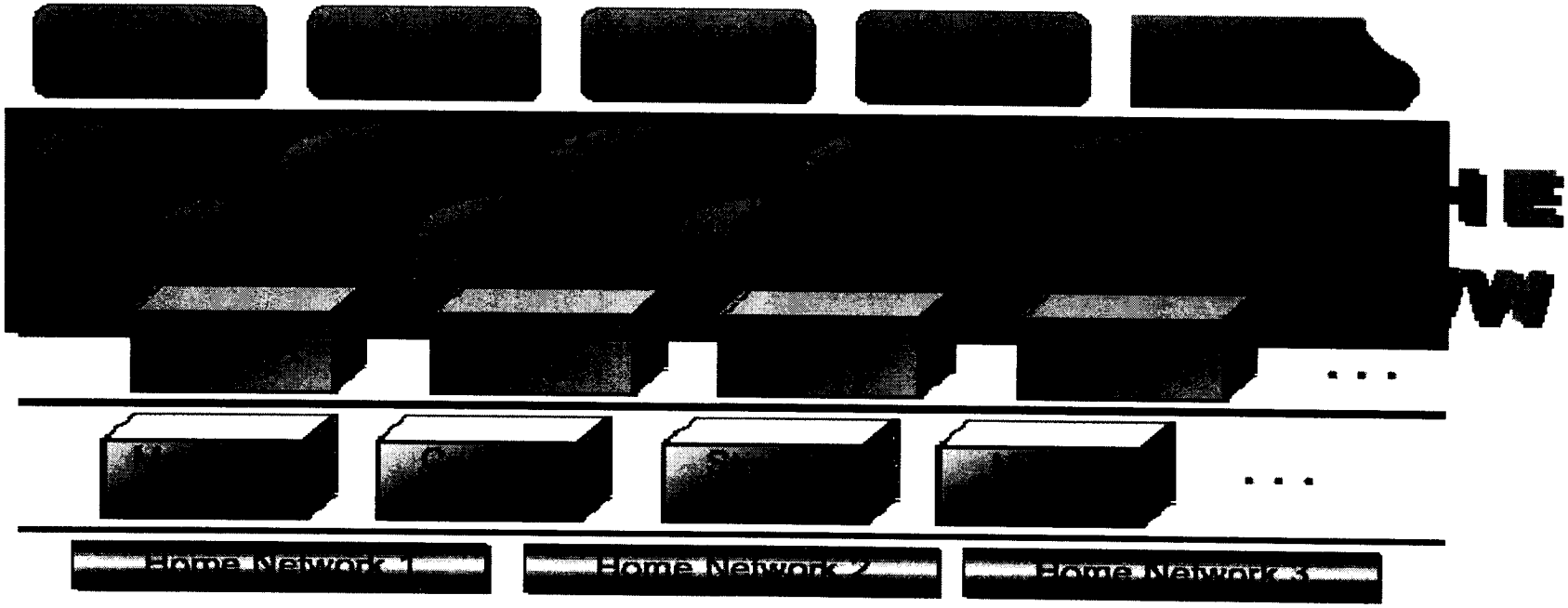
IBM

3. Mobile terminal system requirements

Open System Architecture (OSA)

An Open System Architecture

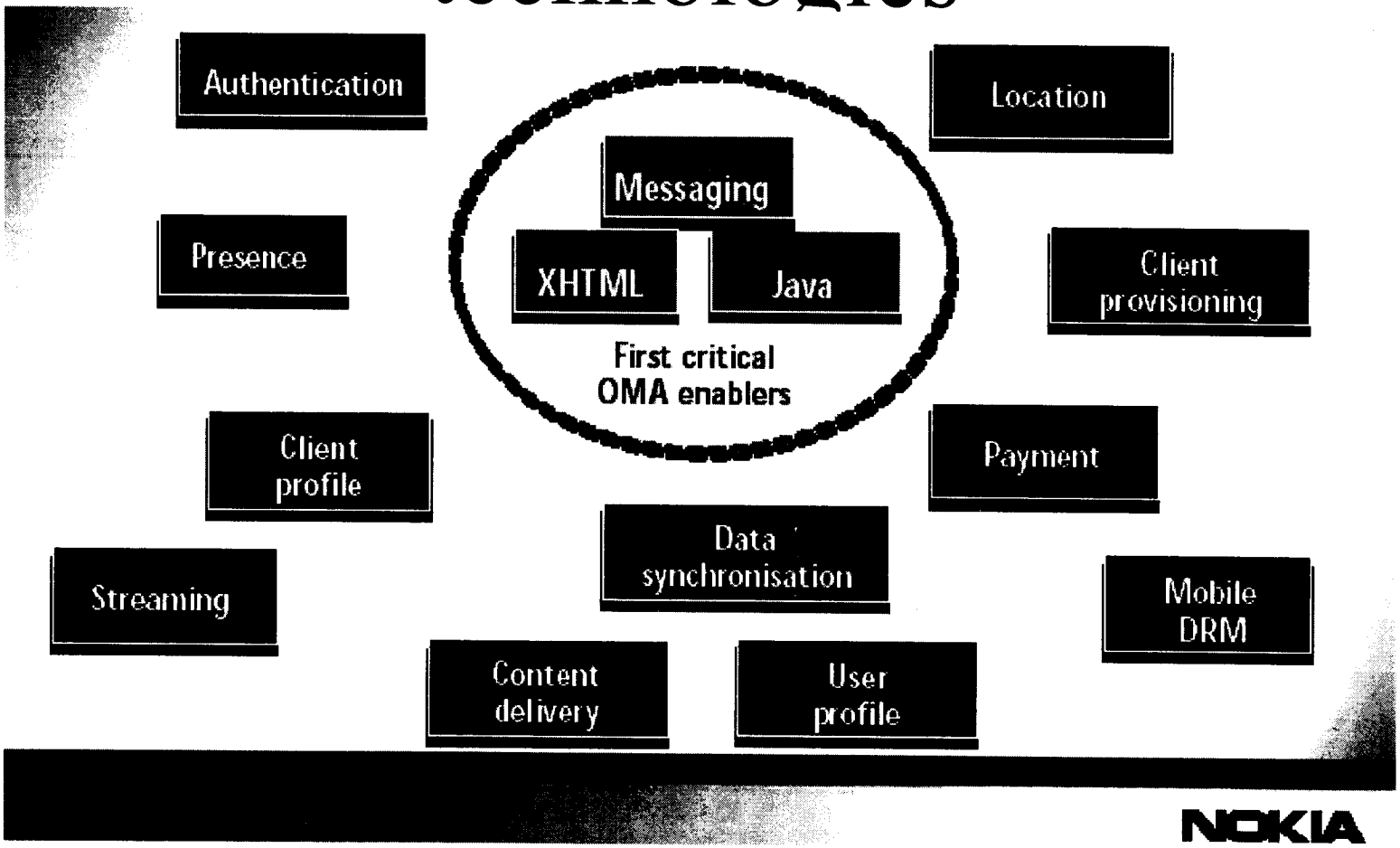
A layered approach in delineating the VHE system architecture Nokia



Open Mobile Architecture (OMA)

- Key elements
  - Open application development environment
  - Open terminal-network and client-server interface
  - Open terminal architecture
  - Open service platform architecture
- Mobile services enablers
  - MMS for messaging
  - XHTML for browsing
  - Java for downloading

Open Mobile Architecture (OMA) technologies



OMA Technology enablers

- Authentication
- Presence
- Client profile
- Streaming
- Data Synchronization
- Content delivery
- User profile
- Mobile DRM
- Payment
- Client provisioning
- Location

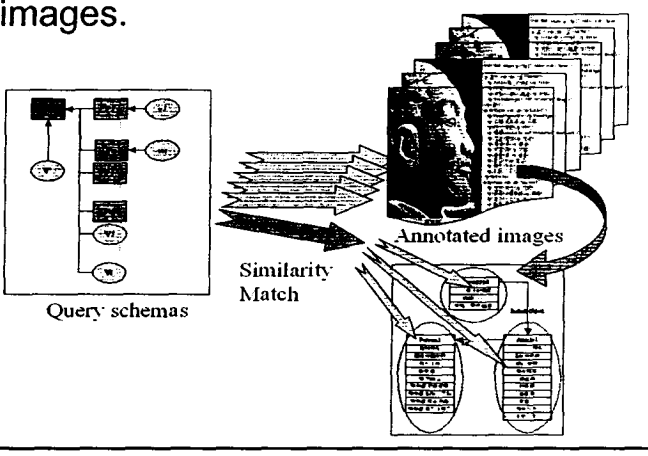
Technologies enabled by III

- Communication Protocol
  - SIP
  - IPv6
- Multimedia CODEC
  - MPEG4
  - JPEG2000
- MMS
- Game engine
- WAP 2.0
- SyncML engine
- J2ME
  - OTA
  - MIDP Next Generation
- VPN

3.Mobile terminal platforms

The Similarity Matching Algorithm

④ Matching a query schema with annotated images.

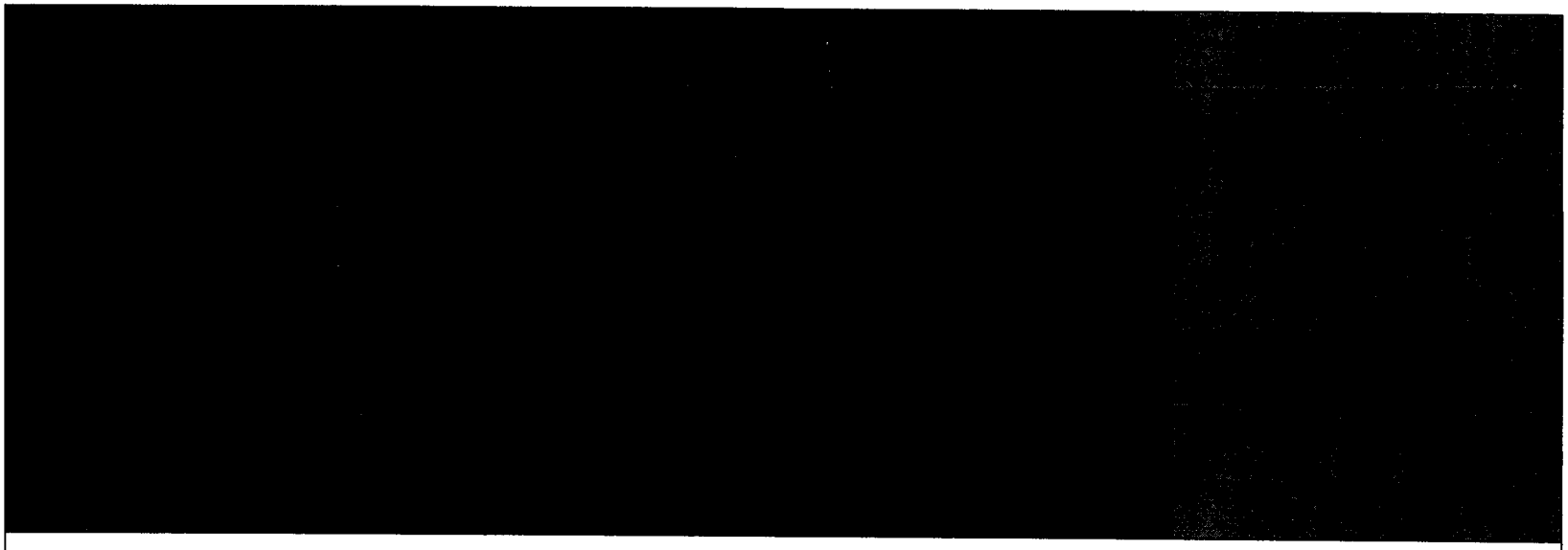


The Similarity Matching Algorithm

④ Method

- Treat the RDF query schema and the RDF query instance as a Tree
- Match all possible interpreting paths of a query instance with annotated pictures.
- Rank the similarity match and find the best answer.

Binary Run Time Environment for Wireless (BREW) QUALCOMM

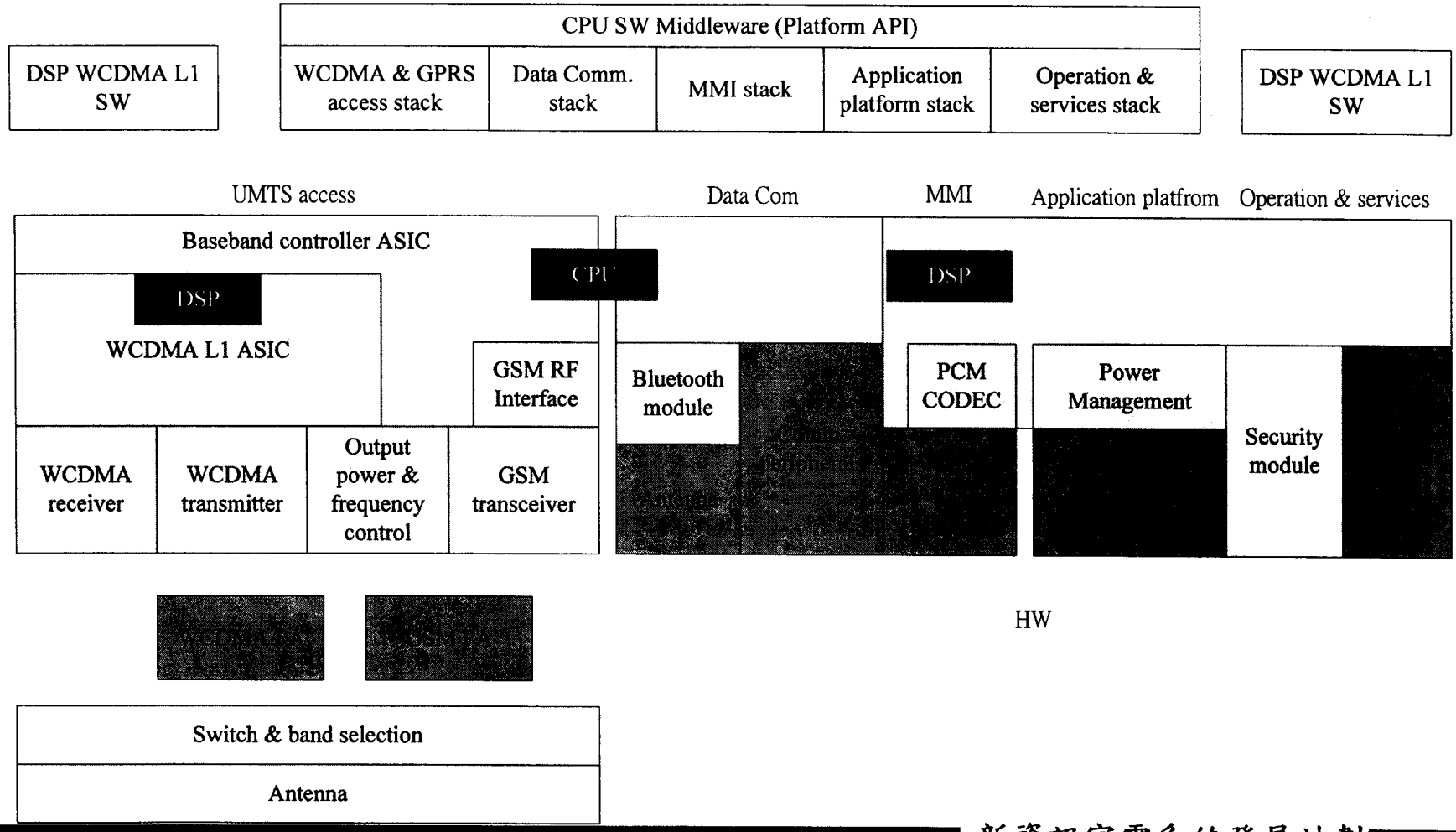


BREW™	
UI Interface	ASIC Software Wireless Internet Launchpad™

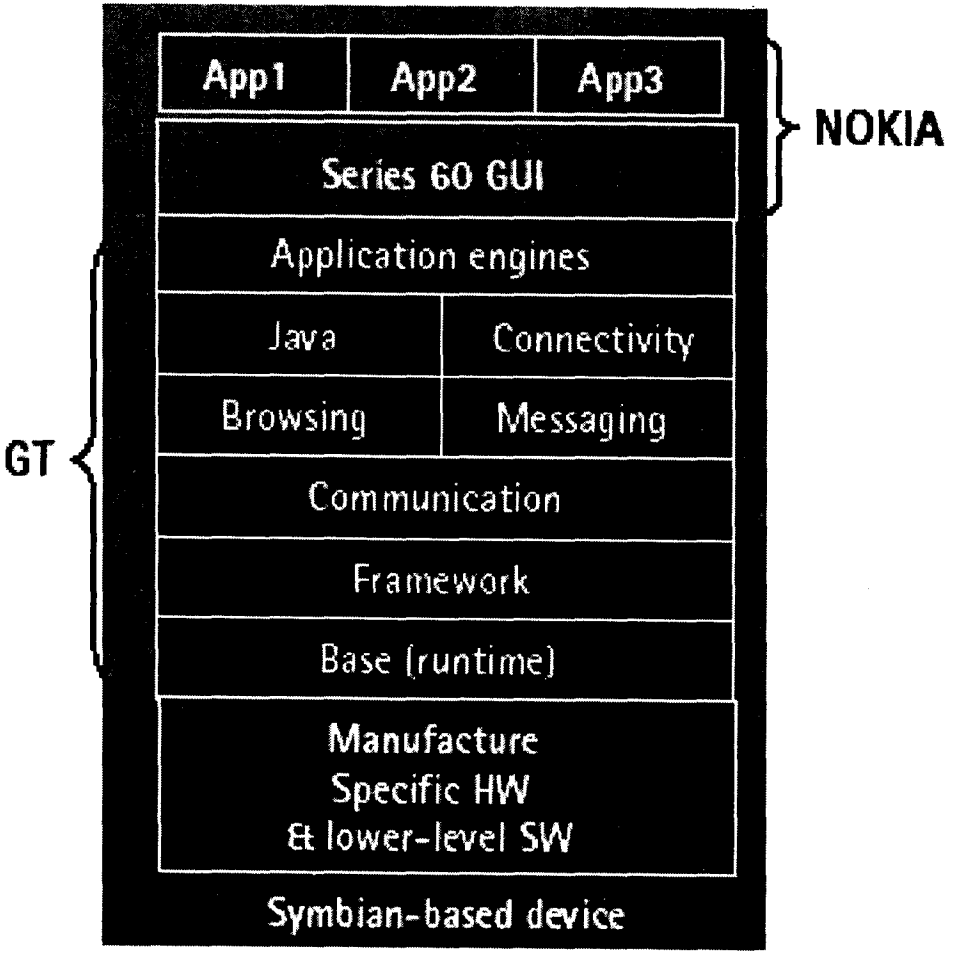
BREW core classes

- Shell services
- Time / timers
- Notifications
- Memory management
- File management
- Database management
- Network and sockets
- HTTP
- Download and billing
- User interface controls
  - Dialogs
  - Date controls
  - Menu controls
  - Text controls
  - Time controls
  - Static text controls
- Display
- Images
- Graphics
- Sound players
- Sound

Ericsson Mobile Platform (EMP) – for UMTS



Nokia Series 60 Architecture

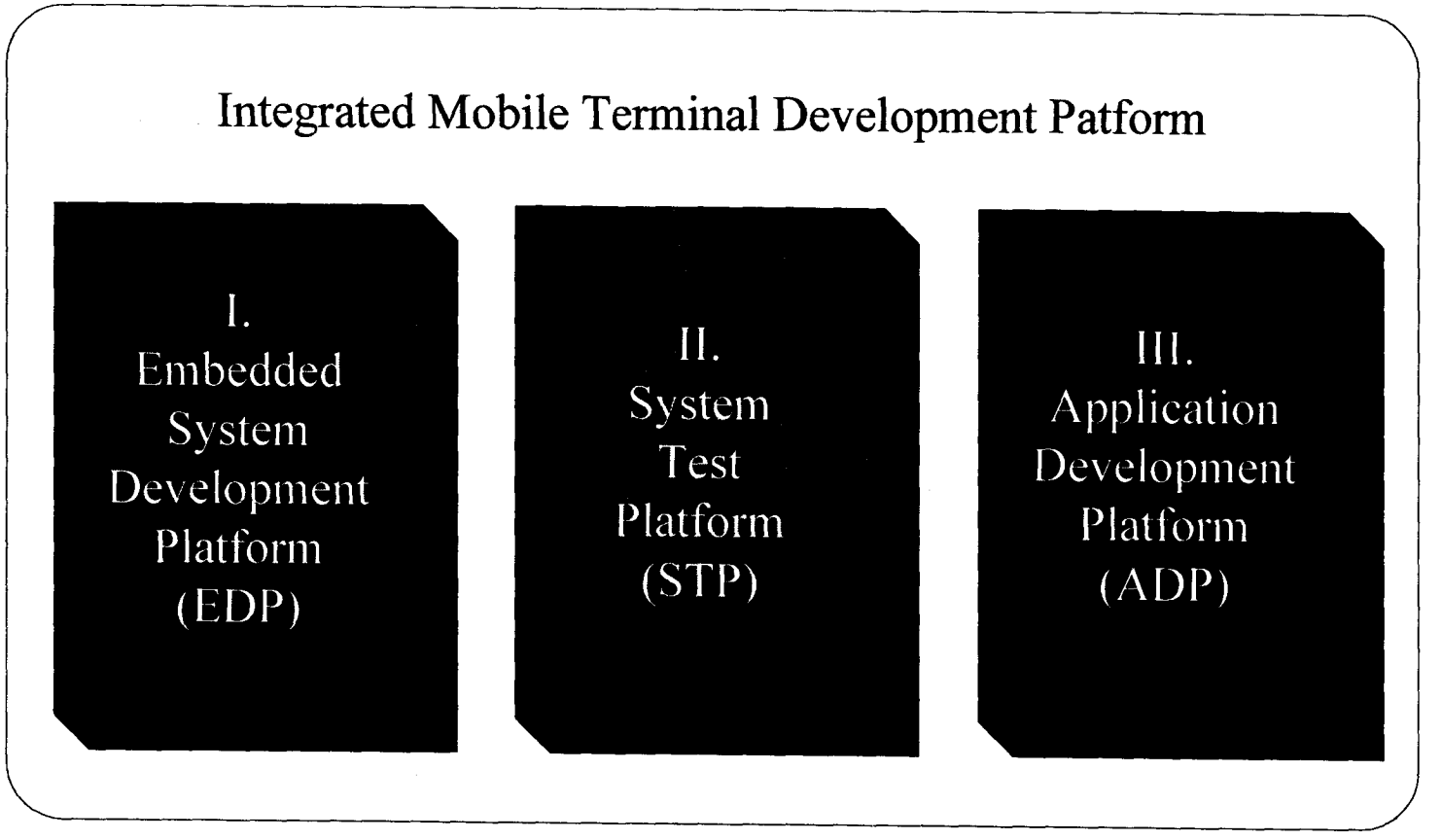


Mobile Terminal Application Development Platform (MAP)

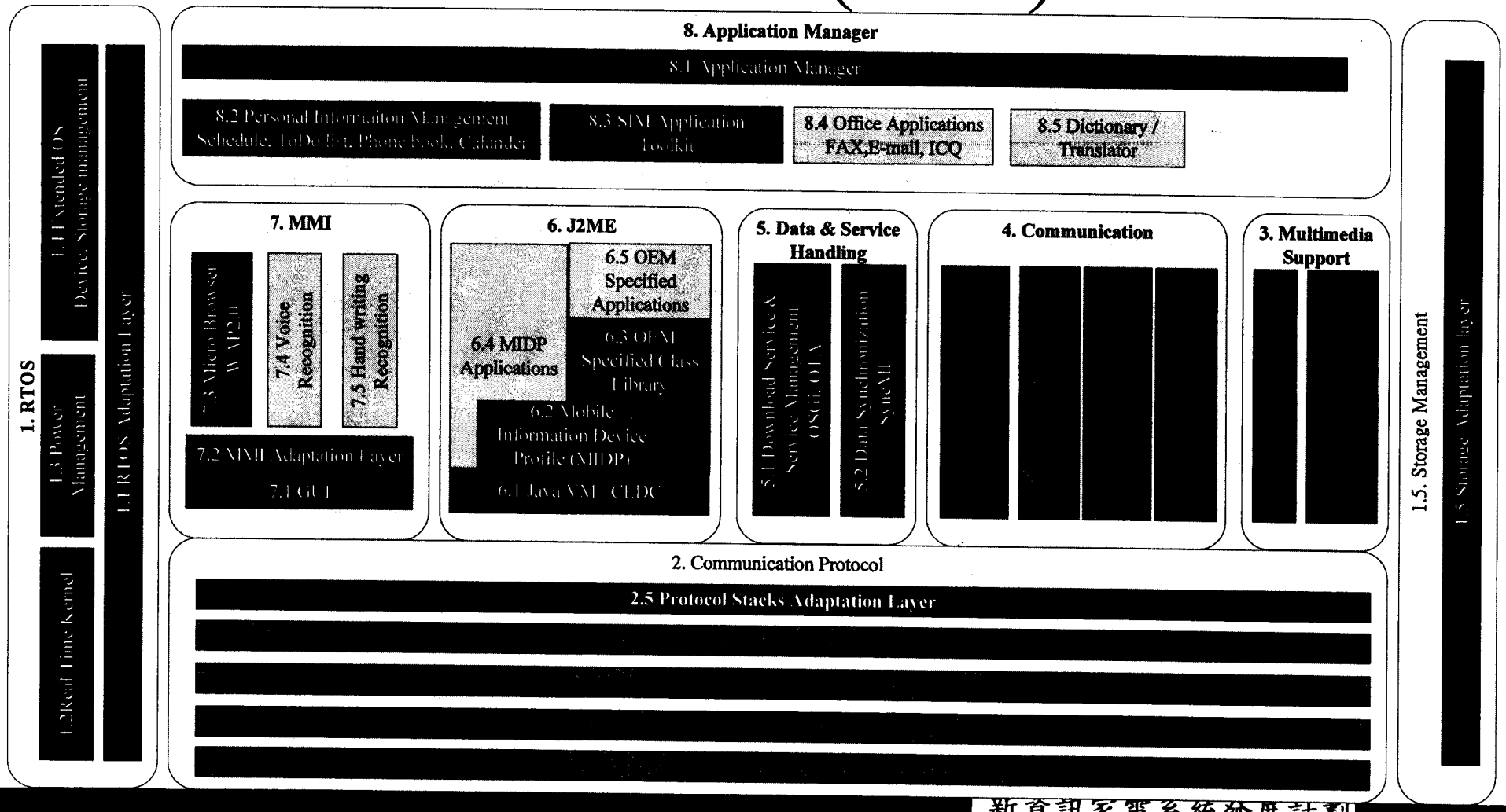
Inventec @ppliances Corp.

Taiwan

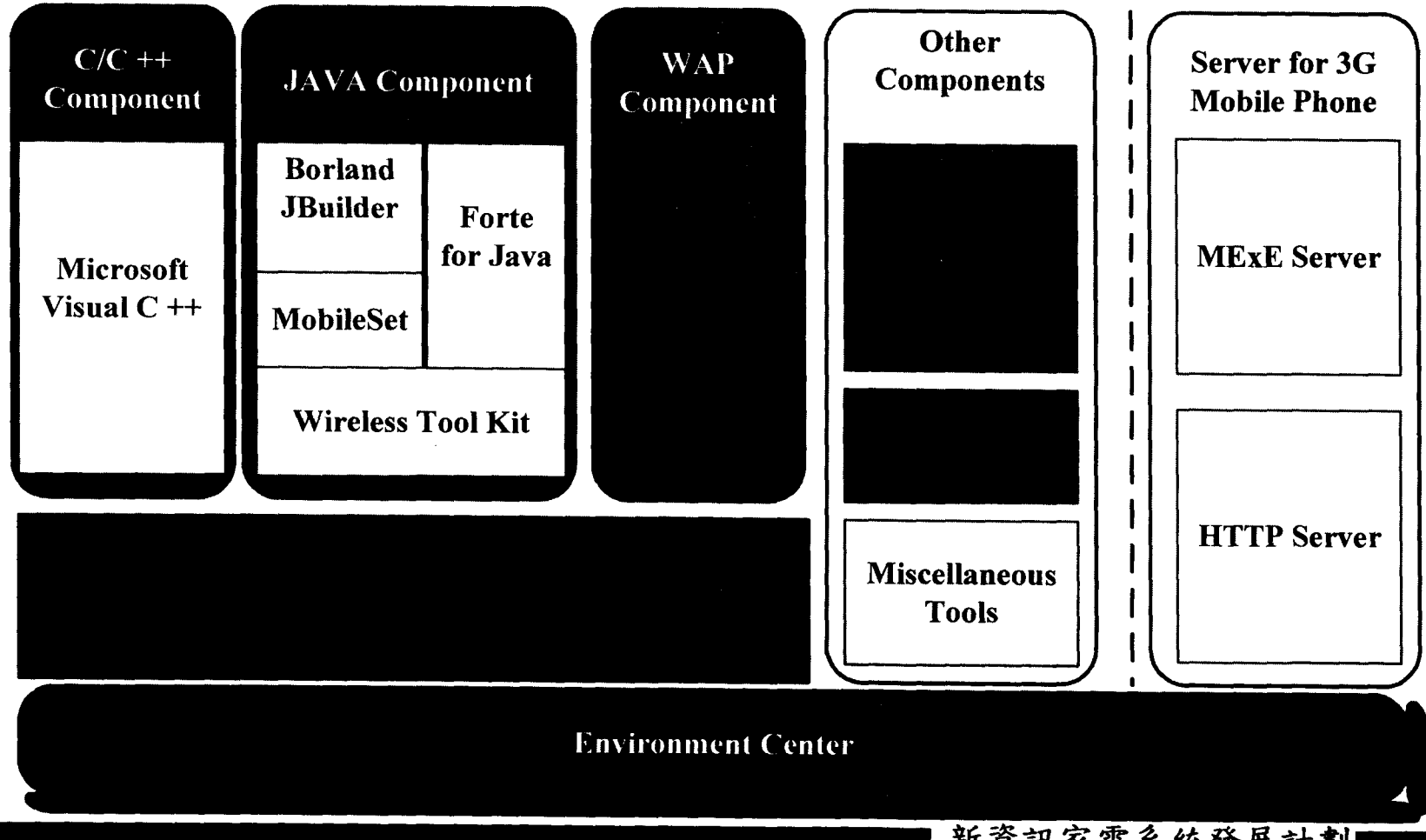
Mobile Terminal Application Development Platform (MAP)



I. Embedded System Development Platform (EDP)



III. Application Development Platform (ADP)



4.Joint development and research for smart terminals

1.經濟部 整合性業界開發產業技術計畫

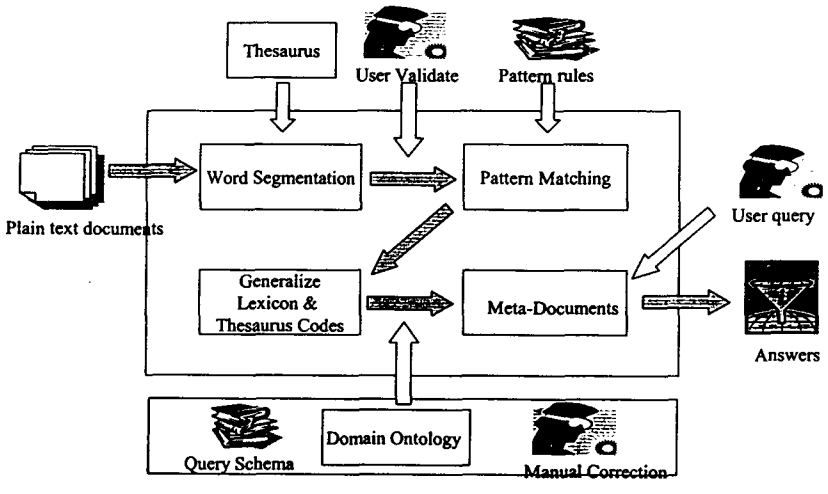
- 行動終端應用發展平台 (Mobile Terminal Application Platform, MAP) 開發計劃
  - 主導廠商：英華達公司
  - 參與廠商：英業達公司，明碁電通，華碩，廣達
  - 時程：
    - 先期研究 1月-4月 2002
    - 全程開發預估 7月2002 – 6月2004

Case Study 2

Answer Simple Historical Questions Using Thesaurus and Ontology

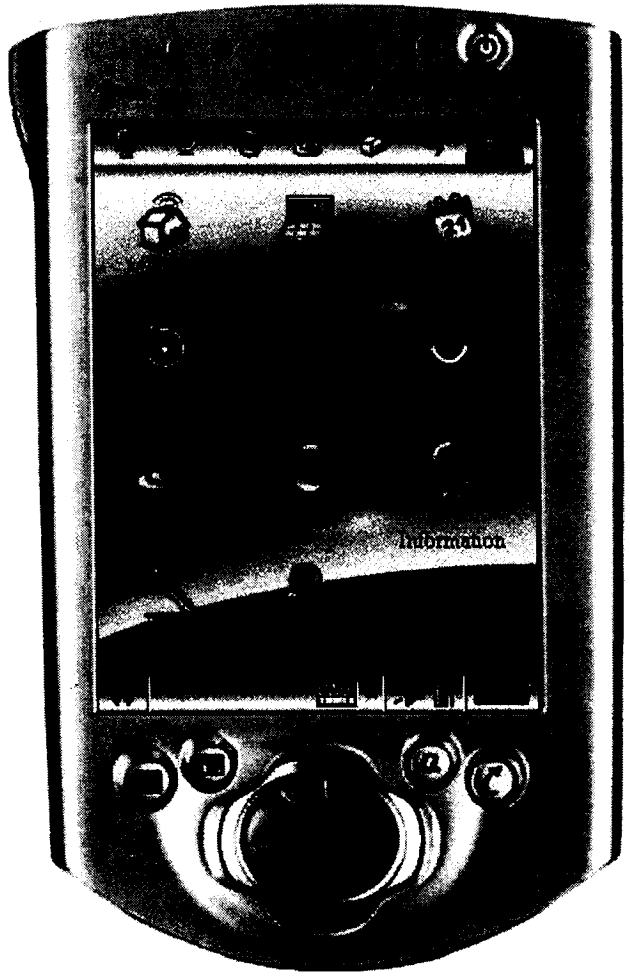


An Ontology-Based Answer Extraction System



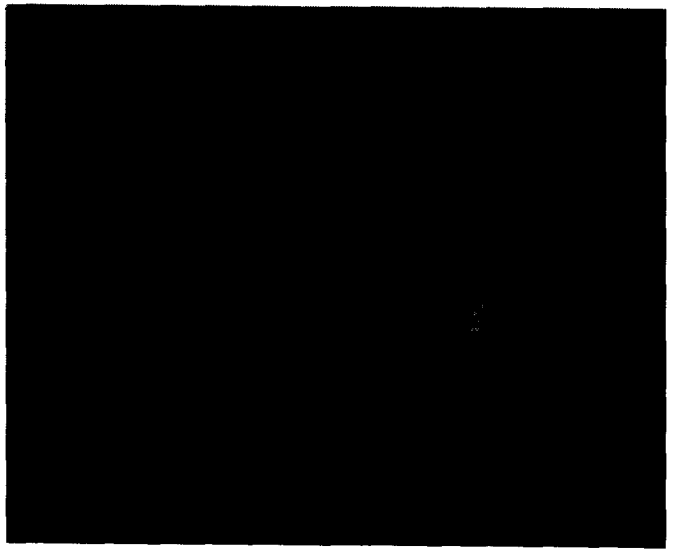
2. IBM及資策會合作開發計劃

- 計劃成果：Wind Pad (掌風)
- 時程：
  - Phase 1: 8月2001- 6月2002
  - Phase 2: x

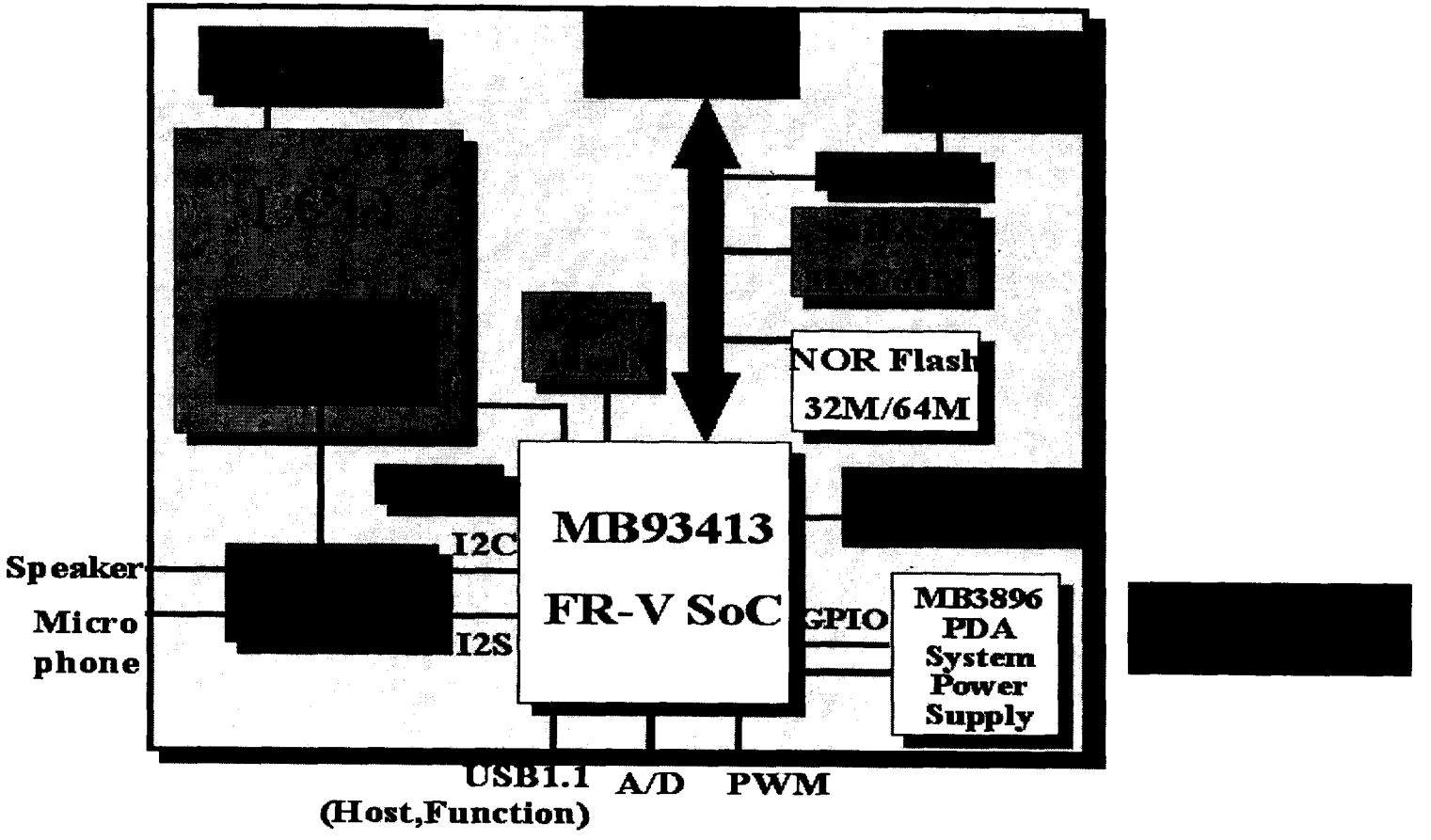


3. Fujitsu & III View Pad project

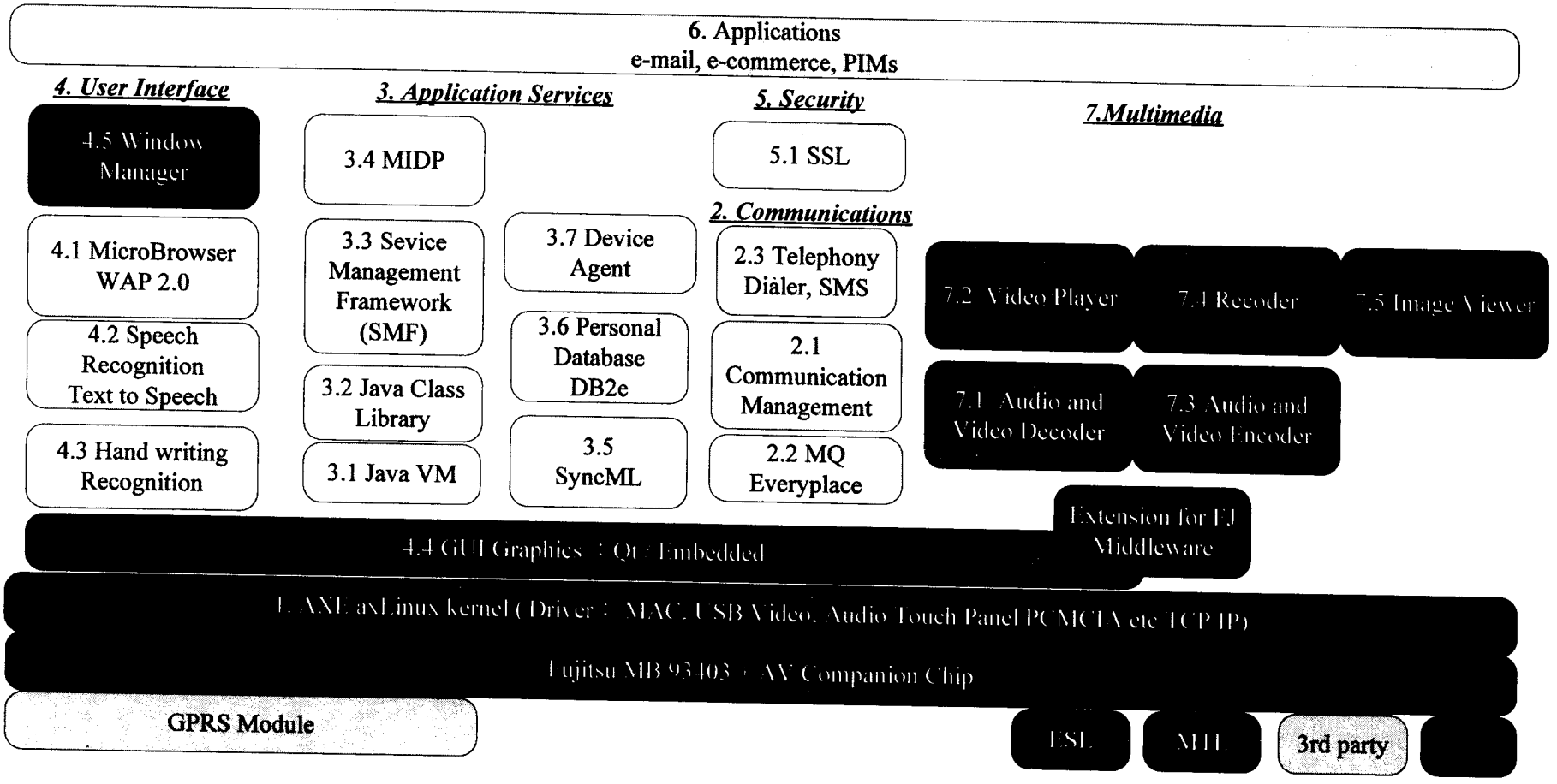
- 時程：6月2002-5月2003
- 開發產品：
  - Media rich PDA solution – FR-V
  - Cell phone solution - MMP
- 合作目的：Devleop World leading media rich smart handheld devices



Fujitsu's media rich PDA solution – FR-V



View Pad software architecture



產學研整合型計劃

- 如何有效提昇產業技術
- 如何有效利用法人之產業技術研發能量
- 如何建立學術研發 Special Interest Group
- 如何有效整合學術研發成果

國科會 Web Technologies 系列研討會(一)

時間：中華民國 91 年 5 月 18 日

會場：臺灣大學資訊工程學系 103 演講廳

主辦單位：國科會 Web Technologies 規畫推動小組

協辦單位：國科會台灣大學資訊工程學系

中央研究院資訊所

議　程

時間	內　容
8:30~9:00	報到
9:00~9:20	致詞 【國科會工程處 蔡忠杓處長】
9:20~10:40	Text and Log Mining on the Web 【中研院資訊所 簡立峰博士】
10:40-10:50	休息
10:50~12:10	Semantic Web 【成大資工系 郭耀煌教授】
12:10~13:10	午餐
13:10~14:30	Web Services Technology and Applications 【資策會 楊仁達博士】
14:30~14:40	Coffee Break
14:40~16:00	電子化政府入口網與各機關系統整合之共通平台初探 【蕃薯藤數位科技協理 張耀仁博士】
16:00~17:00	綜合討論
17:00	散會

聯絡處：台灣大學資訊工程學系
電話：(02) 23625336 轉 303
傳真：(02) 23622704
E-mail：cage@shannon.csie.ntu.edu.tw

目　　錄

⊕ Text and Log Mining on the Web

簡立峰 博士

⊕ Semantic Web

郭耀煌 教授

⊕ Web Services Technology and Applications

楊仁達 博士

⊕ 電子化政府入口網與各機關系統整合之共通平台初探

張耀仁 博士

Text and Log Mining on the Web

簡立峰 博士

中央研究院資訊所

Text and Query Log Mining on the Web

Lee-Feng Chien

Institute of Information Science
Academia Sinica, Taiwan

Outline

I. Web Mining
II. Query Log Mining
III. Text Mining

2

Word segmentation

- ④ It divides the whole document into pieces of lexicons based on Chinese synonym thesaurus.
- ④ It might result in wrong words.

For example,

"將軍政大權集於一身"

Incorrect : "將軍 政 大 權 集 於 一身"

Correct : "將 軍政大權 集 於 一身"

Pattern matching

- ④ It makes complex and continuous fragments into to a unit.

For example,

"13歲"

Original : "1 3 歲"

Result : " 13歲 "

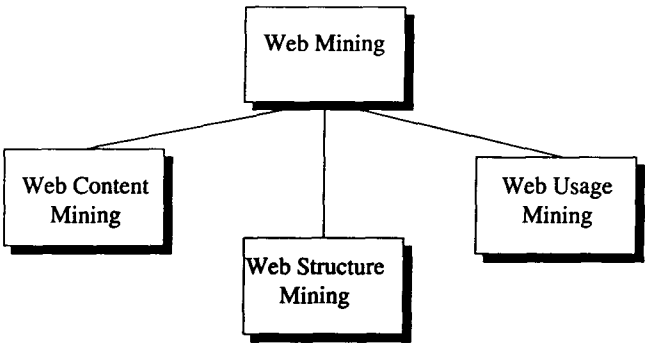
I. Web Mining ?

Web Mining (Srivastava'01)

- **Web is a collection of inter-related files on one or more Web servers**
- **Web Mining**
  - Discovery of interesting patterns from Web content, structure and usage data.
  - A combination of WWW and Data Mining areas (Viewpoint of data mining)
- **Typical Source of Data**
  - Page content
  - Intra-page and inter-page structure
  - Server access logs, Registration information, demographics, past history, etc.
- **Approaches**
  - Agent-based Approach (or AI approach)
  - Database Approach
  - Information Retrieval Approach
  - Information Extraction Approach

4

Taxonomy of Web Mining (Cooley)



5

Discovered Knowledge (DM viewpoint)

- Associations & Correlations
- Sequential Patterns
- Clusters
- Path Analysis
- Others

6

Discovered Knowledge (Web Site Mining)

- Associations & Correlations
  - Page associations from usage/content/structure data
  - Associate rules
- Sequential Patterns
  - Ex: 30% clients who visited /products/software/, had done a search in Yahoo using the keyword "software" before their visit
- Clusters
  - Page clusters, traversal path clusters
- Path Analysis
  - Most frequent paths traversed by users; entry and exit points
- Others

7

Discovered Knowledge (AI/IR/NLP viewpoints)

- Named Entities
- Semantic Templates
- Domain-specific Terms
- Knowledge Bases
- Ontology

8

Web Content Mining

- Most focus on extraction of knowledge from the text of web pages
- Web Page Classification (Chuang & Chien's IRWK'02)
- Text Mining
  - Web Information Extraction
  - XML Mining
  - NLP viewpoint
- Web Image Classification (Tseng's IRWK'02)
- Speech Archive Mining (Chien's ISCSLP'02)

9

Web Page Classification

- The considering problem:

  Automatic categorization of Web hypertext documents into predefined *subject/topic* categories (organized flatly or hierarchically) by using *text and hyperlink information*

- Brief history

  1990-91 **World Wide Web (Nexus) released by Tim Berners-Lee**

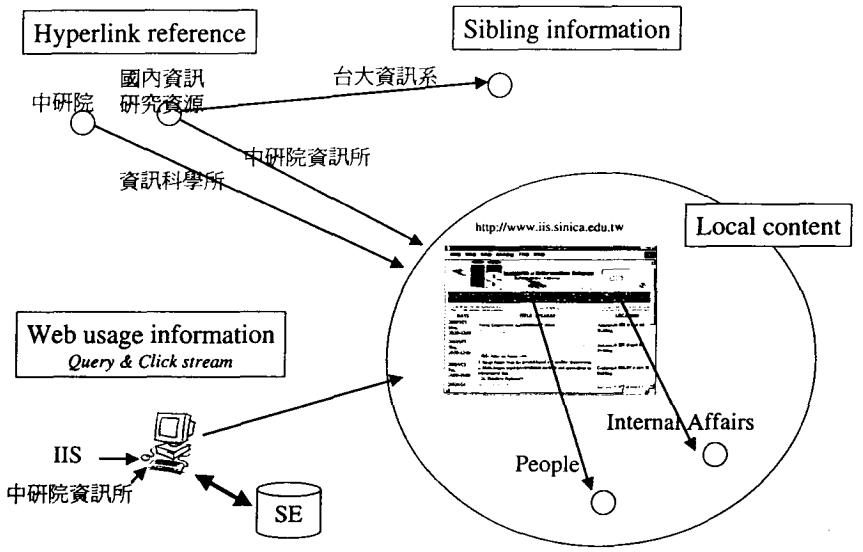
  1993-94 **"Hot list" → Directory Tree, Yahoo!**

  1994-95 **World Wide Web Worm, Altavista**

  1996-200X **Research publications/Projects on the problem of Web page categorization**

10

Hypertext on the Web

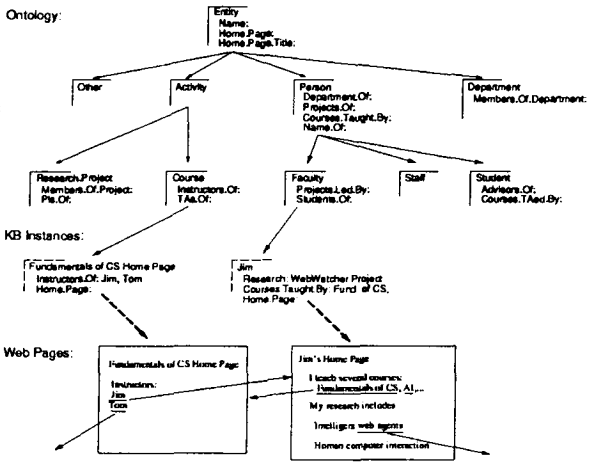


11

Applications

- CMU Web→KB Project (1998-2000) [Craven98]

Classifying Web pages is an essential step to construct Web knowledge base



12

Applications *(cont.)*

- Automatically-constructed, large-scale Web directories
- Web search using automatic classification [Chekuri96]
  - **Class information helps circumvent keyword ambiguity**
- Focused crawling for domain-specific information [Diligenti00]
  - **E.g., CMU Cora (1998)**

13

Text Mining (R. Feldman'95)

- Definition
  - The extraction of implicit (hidden), nontrival previously unknown and potentially useful information from given text data
  - Text data mining, knowledge discovery from textual databases
- First proposal
  - R. Feldman et al., "Knowledge Discovery in Textual Databases (KDT)" in KDD'95.
  - Translate from nonstructure text into traditional database
  - Using a text categorization to annotate text articles with meaningful hierarchical concepts
  - Allowing for interesting data mining operations

14

Text Mining (Mladenic, PKDD'01)

- Text segmentation/summarization
- Topic identification and tracking in time series of documents
- Natural language identification
- Document authorship detection
- Document copying right identification
- Text data visualization
- Automatic text translation
- Question answering
- Speech synthesis

15

Text Mining (M. Hearst, ACL'99)

- TM vs. Information Access
  - Yield tools aid information access, e.g., Create thematic overviews, generate term associations, find general topic and identify central Web pages
- TM vs. Computational Linguistics
  - Help linguistic knowledge acquisition, e.g., augment WordNet relations, extract domain-specific terms, live language modeling , collect bilingual corpus.
- TM vs. Information Extraction ?

16

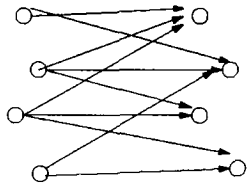
Web Usage Mining

- Data Gathering
  - Web server log, site description data, concept hierarchies
- Data Preparation
  - Distinguish among users, build sessions
- Data Mining
  - Pattern discovery & analysis

17

Web Structure Mining

- Google's Page Rank



hubs authorities

$$PR(A) = (1-d) + d(PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

- Document Citation (siteseer)

18

Semantic Web Mining

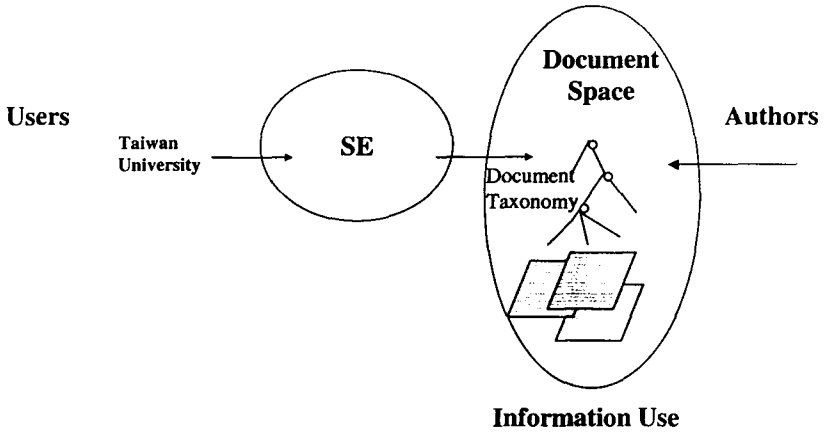
- Current Web
  - Most of Web content is designed for humans to read, not for machine to manipulate meaningfully
- Semantic Web
  - XML+RDF + Ontology + Agent
- Semantic Web Mining
  - Auto-construction of Ontology
  - Case-based reasoning/inference

19

II. Query Log Mining

Short Query

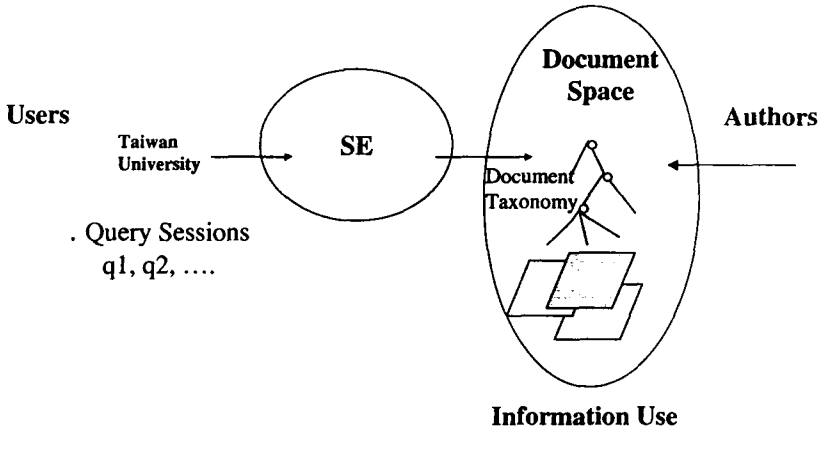
-- is query really short and contains little information ?



21

Short Query

-- rich information may obtain from different types of logs

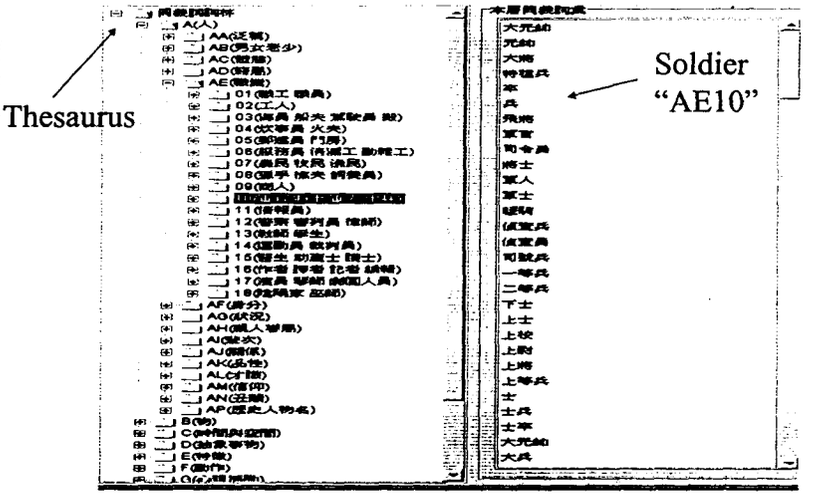


22

Generalization lexicons & thesaurus codes

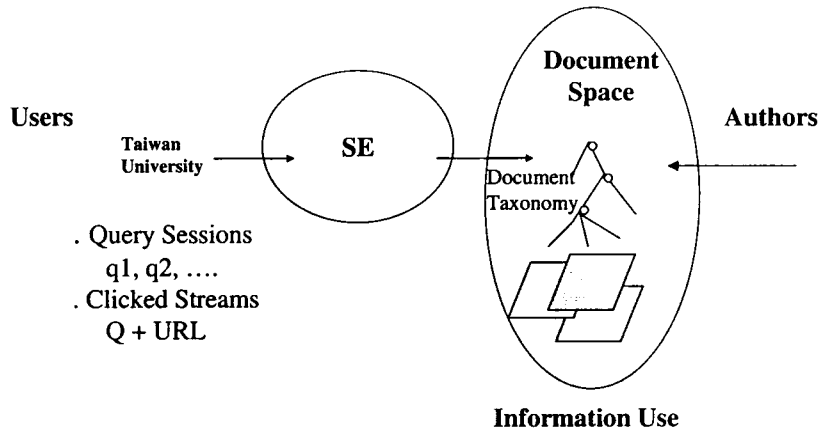
- ④ User may enhance the completeness of the meta-document by domain ontology or linguistic principle.
- ④ Users may also refine the meta-sentence by interacting with an ontology.
- ④ The instance from a meta-document can be expressed in XML/RDF format as knowledge base.

The Chinese Synonym Thesaurus



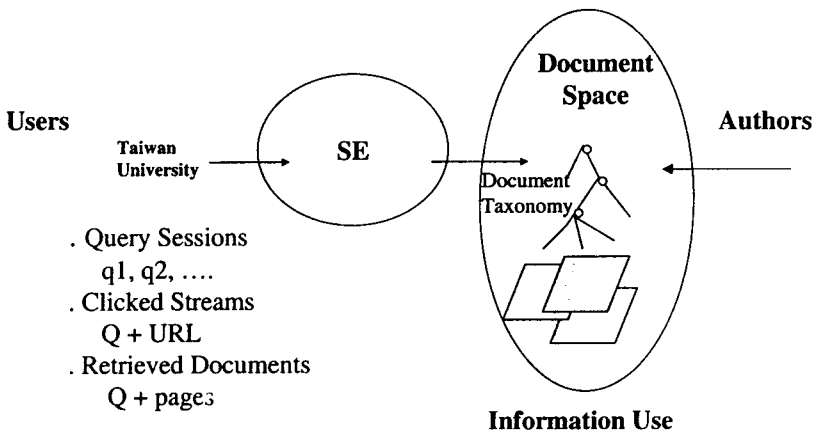
Short Query

-- rich information may obtain from different types of logs



23

Query Log Mining



24

Query Log Mining

- 2M real queries & query sessions
  - Discovery query taxonomy, relevant terms, concepts
  - Develop concept search, personalized search
- Research Topics
  - Query Clustering (ex, exp-file)
  - Query Categorization (DSS'02, ex)
  - Query Session Log Mining (WI'01, WWW'01, ex)
  - Taxonomy Generation (ex)

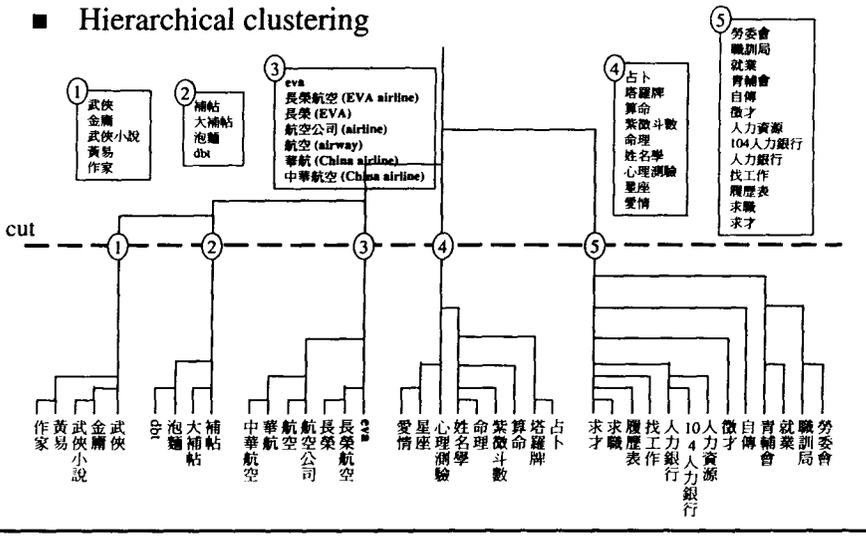
Next

25

Back

Query Clustering

- Hierarchical clustering



26

An Illustrative Example

nude → Google

Creative Nude Photography Network -- Fine Art Nude and ...
... The Creative Nude and **Erotic Photography** Network is the number one net portal to the best in fine art nude and **erotic photography**! Over 100 CNPN Member Sites ...

Nude Places
... to be **naked**. Walking in the forest, cruising the lake in open boats, swimming, picnicking and nude photography are all enjoyed in the nude. 60 minutes $39.95. ...

A Brave Nude World
... A Brave Nude World! Warning: This site contains links to fine art nude & **erotic photography**. If you are under 18 or do not wish to view this material, You can ...

Co-occurred feature terms

term	tf/df	class
erotic photography	3/2	Adult
naked	1/1	Adult
photography	2/2	Art
art	3/2	Art
...	...	...

27

Extraction of Basic Feature Terms

- **Performance of different features: randomly selected, hi-frequency, and *seed terms***
  - Popular queries not affected by ephemeral trends, e.g., "movie", "basketball", "mutual fund", etc.
  - More expressive and *distinguishable* in describing a particular category
  - Two logs compared and extracted *9,709* overlapping top query terms as feature terms

G-1999 / D-1998	Top 1,000 terms	top 20,000 terms	ALL
Top 1,000 terms	583/58.30%	977/97.70%	992/99.20%
Top 20,000 terms	914/91.40%	9,709/50.71%	14,721/76.89%

28

Hierarchical Term Clustering

- **Term Similarity**
  - Cosine measurement
- **Agglomerative hierarchical clustering**
  - Compute the similarity between all pairs of clusters
    - Estimate similarity between all pairs of composed terms
  - Merge the most similar (closest) two clusters
    - Single link method
    - Complete link method
  - Update the cluster vector of the new cluster
  - Repeat steps 2 and 3 until only a single cluster remains

29

Preliminary Experiment

- **Query Clustering**
  - Test queries
    - Top 1k queries, random 1k queries
  - Level of categories
    - 10 first-level, 100 second-level
  - Feature set
    - Top 1k core terms, 2k core terms, ...
  - Evaluation metrics
    - F-Measure

30

Evaluation: F-Measure

query. We then calculate the recall and precision of that cluster for each given class. More specifically, for cluster j and class i

Recall(i, j) = n_{ij} / n_i

Precision(i, j) = n_{ij} / n_j

where n_{ij} is the number of members of class i in cluster j, n_j is the number of members of cluster j and n_i is the number of members of class i.

The F measure of cluster j and class i is then given by

F(i, j) = (2 * Recall(i, j) * Precision(i, j)) / ((Precision(i, j) + Recall(i, j))

For an entire hierarchical clustering the F measure of any class is the maximum value it attains at any node in the tree and an overall value for the F measure is computed by taking the weighted average of all values for the F measure as given by the following.

$$F = \sum_i \frac{n_i}{n} \max\{F(i, j)\}$$

where the max is taken over all clusters at all levels, and n is the number of documents.

31

Preliminary Results

*** F- measure**

- For top 1000 queries: 0.4767
- For random 1000 queries: 0.3965

*** Variant feature sets (top-N core terms as features)**

[redacted]

32

Back Next

Query Categorization

Query term (translation in English)	Categories assigned by humans	Top 5 categories suggested by the machine					Major category	/ Subcategory	:Symbol
chinatrust (a commercial bank)	bb	bm	tl	tf	*bb*	cn	Arts & Humanities	/ Arts	: aa
							Business & Finance	/ Banks	: bb
wap	bn	lp	cn	bb	bm	kb		/ Electronics Industry	: be
								/Business Information	: bf
real player		cd	cn	mn	ch	pc		/ Personal Finance	: bm
								/ Telecom Industry	: bn
michael jackson			cd	tl	en	cn	Computers & Networks	/ Software	: cd
								/ Hardware	: ch
								/ Network Services	: cn
								/ Search Engines	: cs
								/ Local Culture	: pc
sheng-dan-ka (christmas card)	lo	*lo*	aa	be	fs	cs	Politics & Society Adult	/ Sex Photos	: sp
								/ Sex Info.	: ss
xie-zhen-ji (sex-related photos)	sp	*sp*	cn	cs	ss	cm	Travel	/ Travel Abroad	: tf
								/ Travel Local	: tl

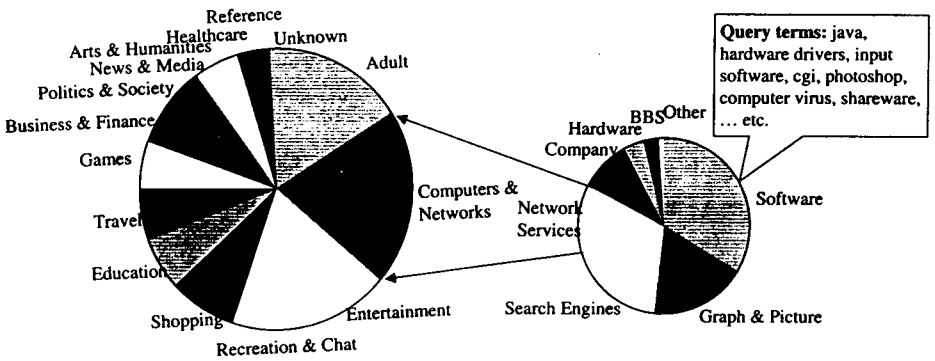
"chinatrust" (a commercial bank in Taiwan)

Assigned by human: *bb (bank)*

Suggested by machine: bm (personal finance), tl (travel-local), tf (travel-abroad), *bb (bank)*, cn (network service)

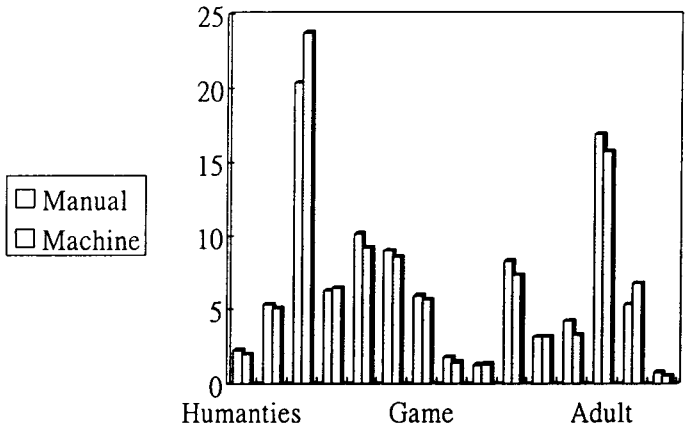
33

Distributions of Search Interests in Various Categories (Pu, JASIS'02)



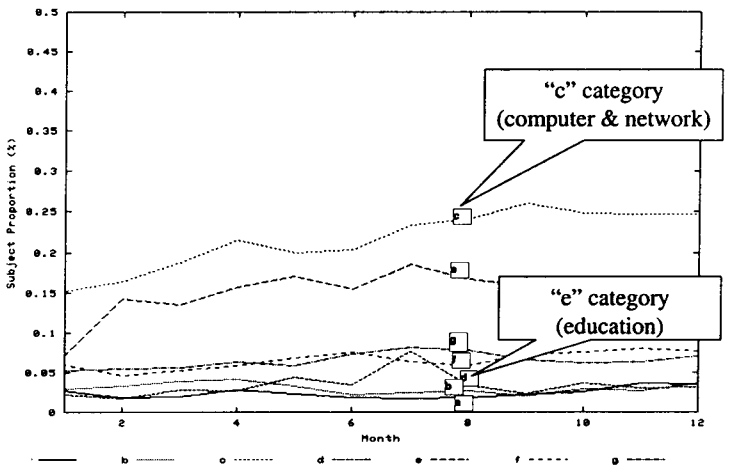
34

Comparison between Distributions of Manual and Auto categorization



35

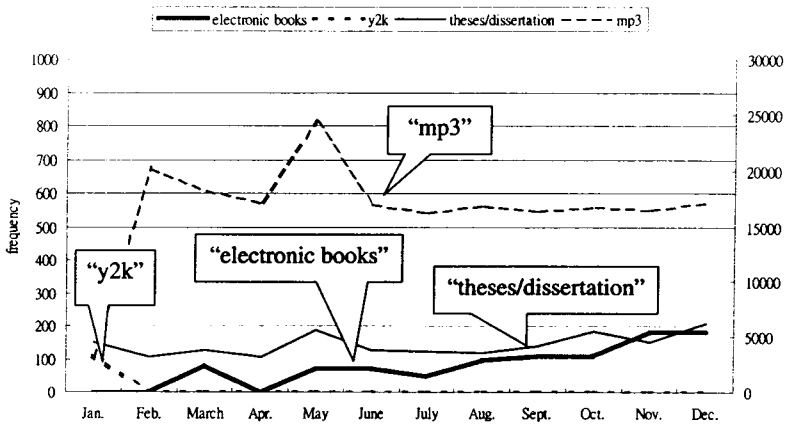
Trends of Search Interests (by Categories)



36

Back

Trends of Search Interests (by Query Term)



37

Relevant Term Extraction

Query-Session-Based	Document-Based
新浪網 (Sina)	新浪網 (Sina)
pchome	著作權 (copy right)
yahoo	智慧財產權 (IP right)
kimo	聊天 (chat)
sex	新聞 (news)
yam	nasdaq
雅虎 (Yahoo)	聊天室(chat room)
tomail	理財 (personal finance)
pc+home	金融 (banking)
免費信箱 (free email)	華淵 (company name)
中文雅虎(Yahoo Chinese)	信箱 (mail box)
奇摩 (Kimo)	翻譯 (translation)
搜尋引擎 (search engine)	算命(fortune telling)

38

Observation on Query Session

- **Example**
  - **[ID1]** 台大醫院, 台大醫院公館院區, 台大醫院公館區
  - **[ID2]** 台大醫院, 台大醫學院, 台大醫學院圖書館, 醫學期刊
  - **[ID3]** 台大醫院, 台大兒童醫院, 長庚兒童醫院
  - **[ID4]** 台大醫院, 台大醫院掛號, 台大醫院網路掛號
  - **[ID5]** 台大醫院, 三軍總醫院, 八一七醫院, 萬芳醫院
  - **[ID6]** 台大醫院, 臺大醫院, 台大, 臺大

39

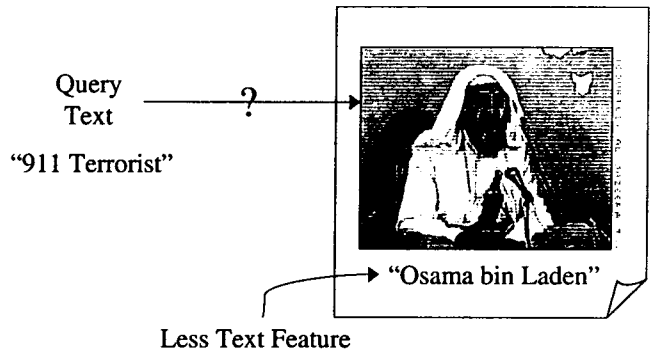
Comparison with Document-based Approach

	Precision	Recall	Comments
Document-Based	0.25	0.66	Low Precision Useful for low-frequency queries
The Approach	**0.90**	**0.45**	High Precision Low Recall More Comprehensive Need a lot of logs

40

Query Expansion for Web Multimedia Search (PCM'01)

- There is a large gap between query text and target image contents, while using the pattern-match search and surrounding text index.

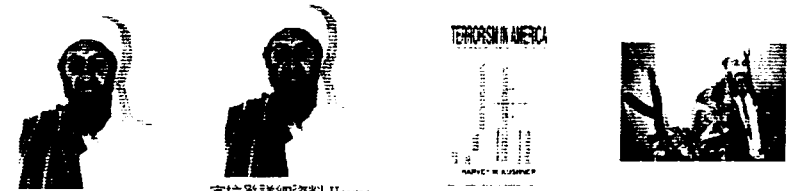


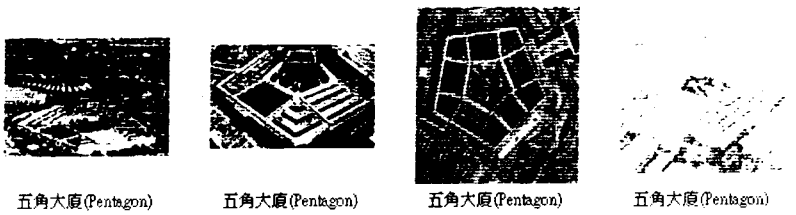
41

Back

Example of Session-Based Query Expansion

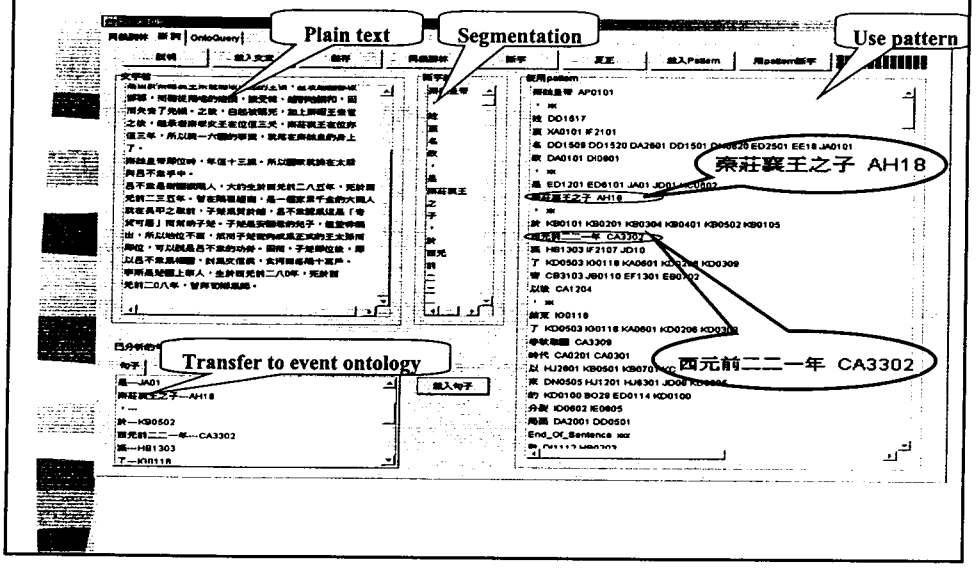
Relevant terms: 世貿大樓, 911, 五角大廈, 恐怖主義, 賓拉登, 波音



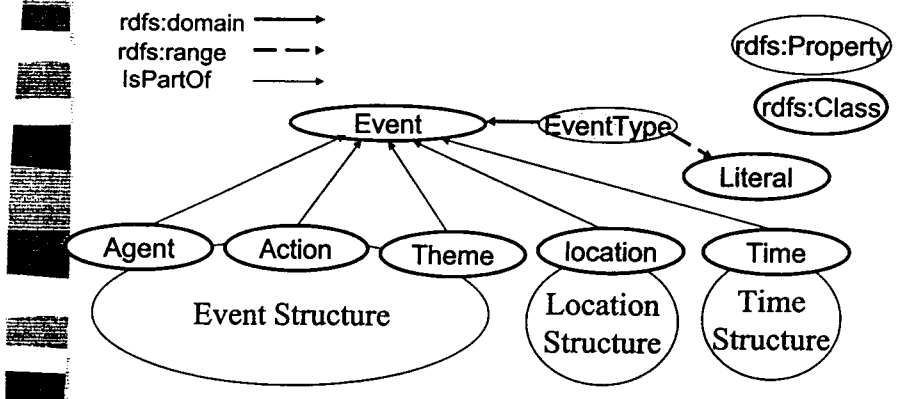


42

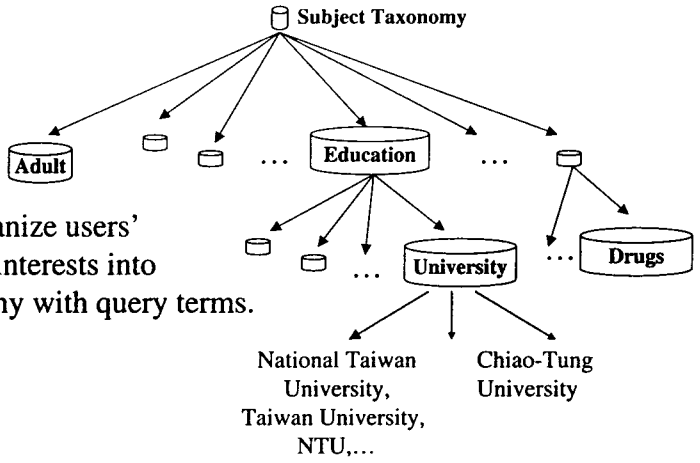
Word Segmentation Post Editing Tool



Event Ontology



Query Taxonomy

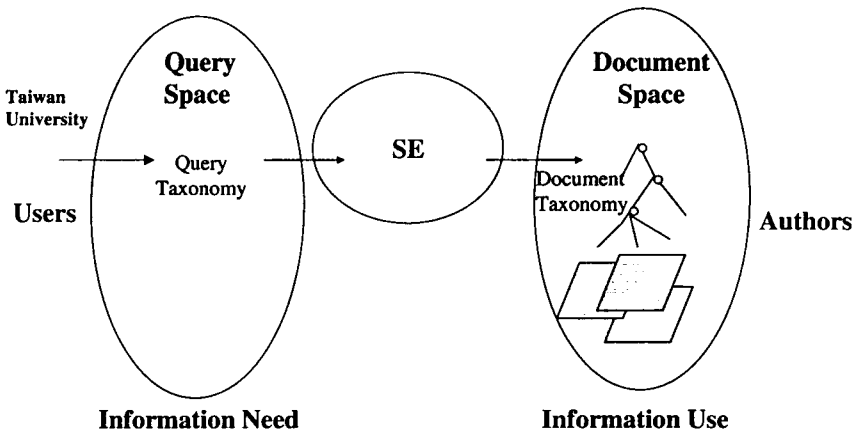


To organize users' search interests into hierarchy with query terms.

43

Query Space

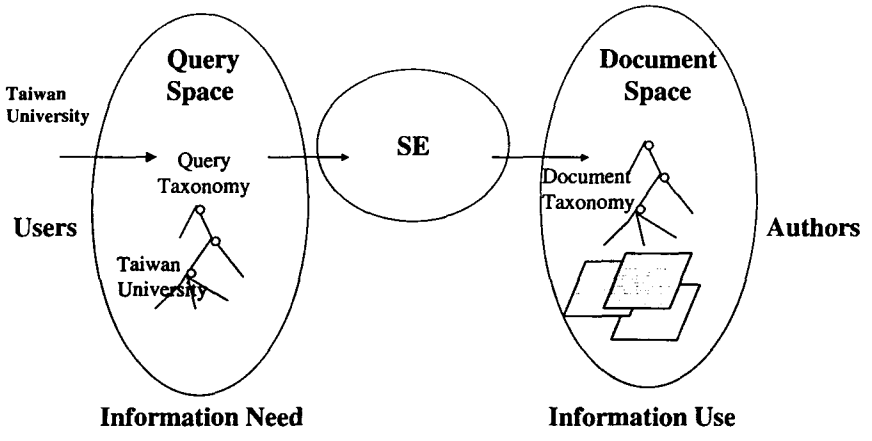
-- less investigated; can show popular requests, ...



44

Query Space

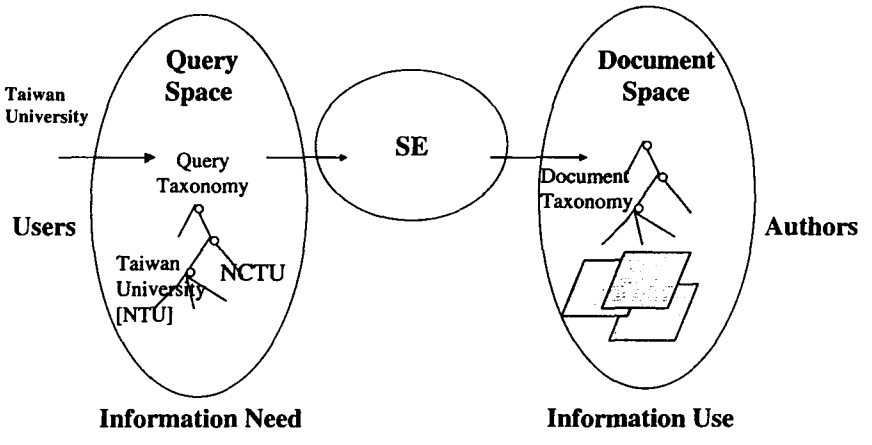
-- may contain related requests & alternative terms



45

Query Space

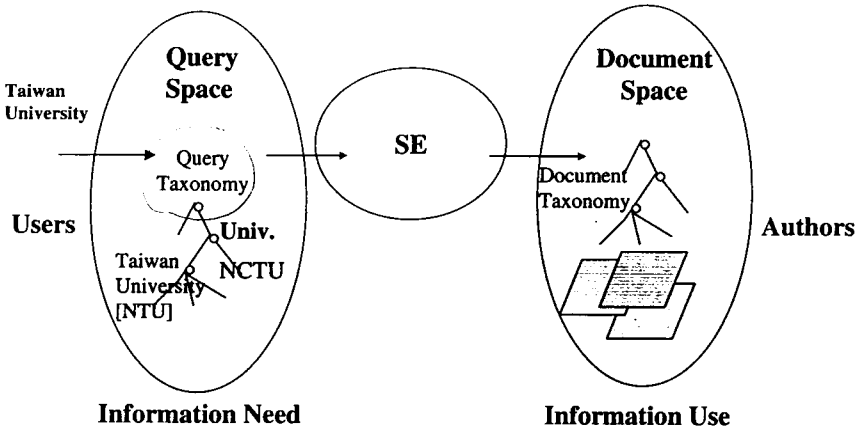
-- may contain related requests & alternative terms



46

Query Space

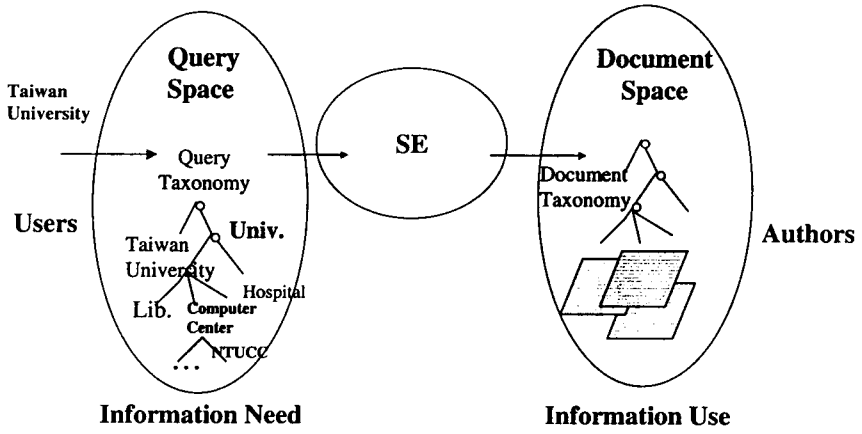
-- may form classes of similar queries



47

Query Space

-- contain related requests & alternative terms

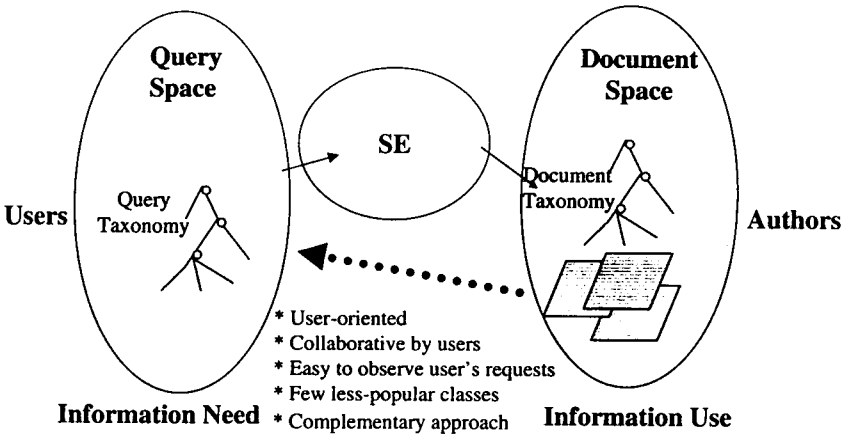


48

Back

Document Classification

-- based on query taxonomy



49

III. Text Mining

Text Mining

- 2M Web pages & anchor texts
  - **Discovery domain-specific terms, query translations**
  - **Develop Cross-language search, multilingual classification**
- Research Problems
  - **Anchor Text Mining**
    - Translation Extraction of Query Terms (ICDM'01, ACM TALIP'01, exp file, ex)
    - Transitive Translation using Third Language (COLING'02)
  - **PAT-tree-based Text Mining**
    - Chinese Term Extraction (SIGIR'97, IP&M'00)
    - Domain-specific & Incremental Term Extraction (Advances in Computational Terminology'02)
    - Incremental Classification (PAKDD'02)

Next

51

Anchor Text Mining for Query Translation

Anchor Text

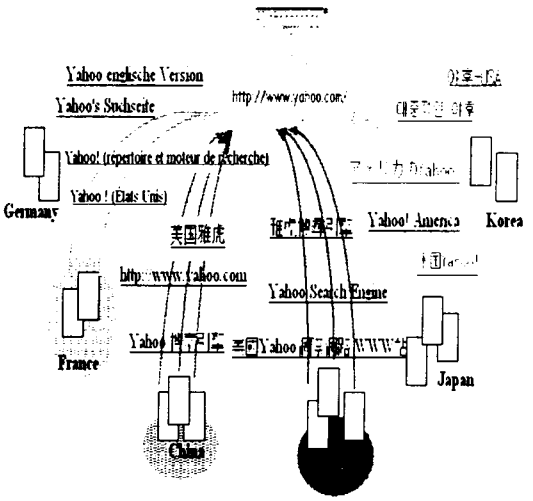
-- the descriptive part of a link of a Web page

-- contains some valuable information

- ***multi-lingual terms***
- **synonym**
- **abbreviation (acronym)**
- **domain-specific terms**
- **topic concept**
- **URL**

Anchor-Text Set

-- a set of anchor texts linking to the same page



52

Query Translation

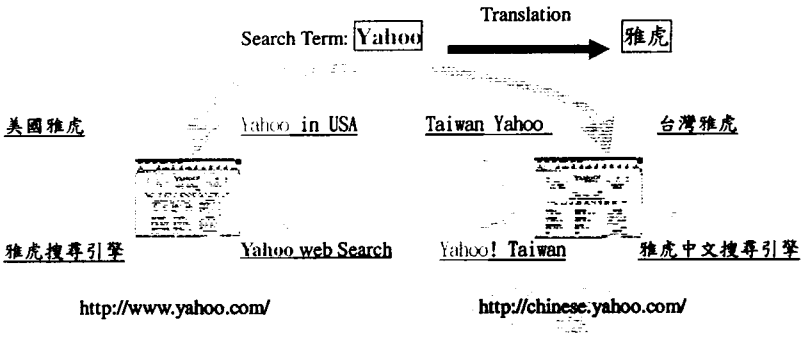


Fig.1. Term translation encompassed in anchor text

53

Probabilistic Inference Model

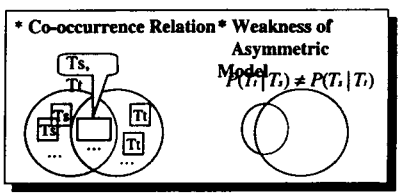
- Asymmetric model: $P(T_t|T_s) = \frac{P(T_s \cap T_t)}{P(T_s)}$
- *Symmetric model with link information* :

$$P(T_s \leftrightarrow T_t) = \frac{P(T_s \cap T_t)}{P(T_s \cup T_t)}$$

$$= \frac{\Sigma_{U_i} P(T_s \cap T_t|U_i)P(U_i)}{\Sigma_{U_i} P(T_s \cup T_t|U_i)P(U_i)}$$

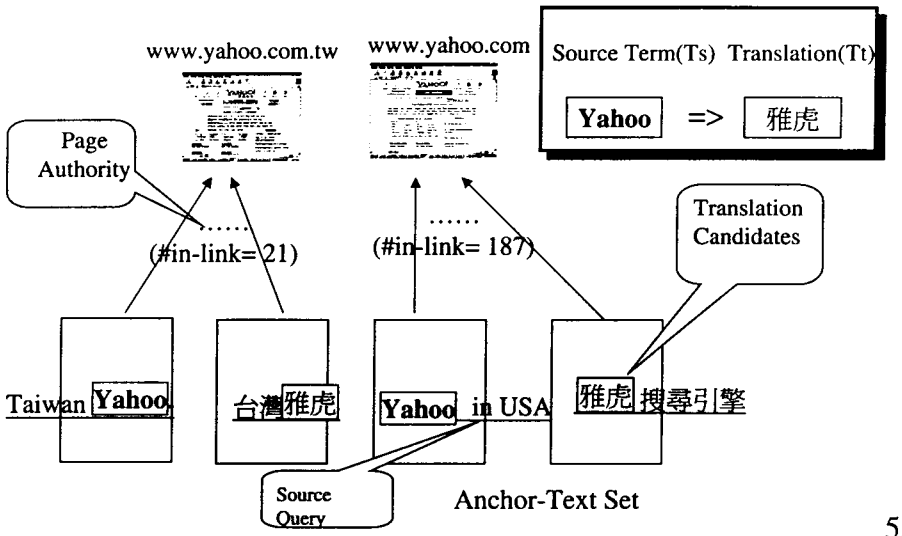
$$\approx \frac{\Sigma_{U_i} P(T_s|U_i)P(T_t|U_i)P(U_i)}{\Sigma_{U_i}[P(T_s|U_i) + P(T_t|U_i) - P(T_s|U_i)P(T_t|U_i)]P(U_i)}$$

where $P(U_i) = \frac{\#\textit{in-link of } U_i}{\#\textit{ total in-link}}$



54

Observation of Anchor-Text



55

Back

Example

English	Traditional Chinese	Simplified Chinese	English	Traditional Chinese	Simplified Chinese
Yahoo	雅虎	雅虎	Internet	網際網路	互联网
Intel	英代爾	英特尔	network	網路	网络
Dell	戴爾	戴尔	online	線上	在线
HP	惠普	惠普	download	下載	下载
Nokia	諾基亞	诺基亚	homepage	首頁	主页
Ericsson	易利信	爱立信	computer	電腦	计算器
Sony	新力	索尼	hardware	硬體	硬件
Nike	耐吉	耐克	database	資料庫	数据库
Stanford	史丹佛	斯坦福	information	資訊	信息
Sydney	雪梨	悉尼	game	遊戲	游戏
White House	白宮	白宫	movie	電影	电影
search engine	搜尋引擎	搜索引擎	news	新聞	新闻
Star War	星際大戰	星球大战	video	影視	录像
police	警察	公安			

56

Performance

Top-n inclusion rates obtained with three different approaches.

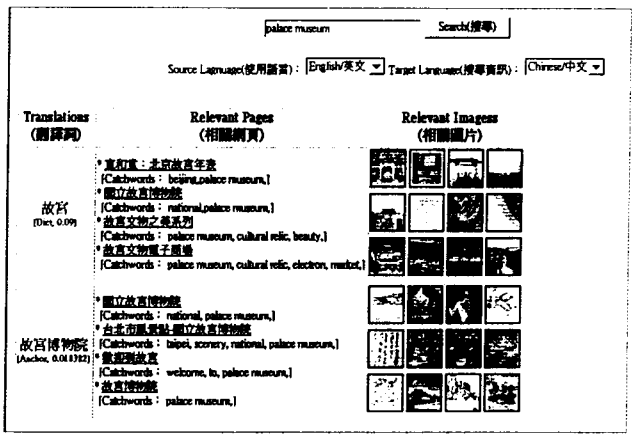
Approaches	Top1	Top2	Top3	Top4	Top5
Anchor Text Mining	57.0%	68.6%	74.3%	77.9%	80.1%
Dictionary Lookup	30.5%	30.5%	30.5%	30.5%	30.5%
Combine Anchor Text Mining and Dictionary	74.0%	82.9%	86.7%	88.9%	90.2%

57

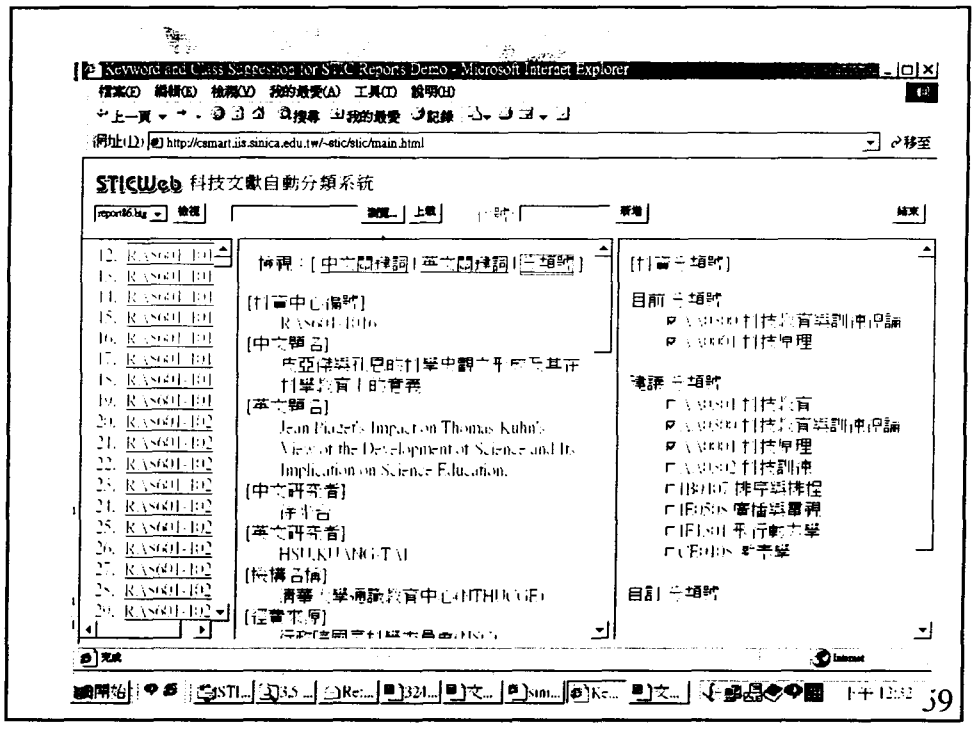
Back

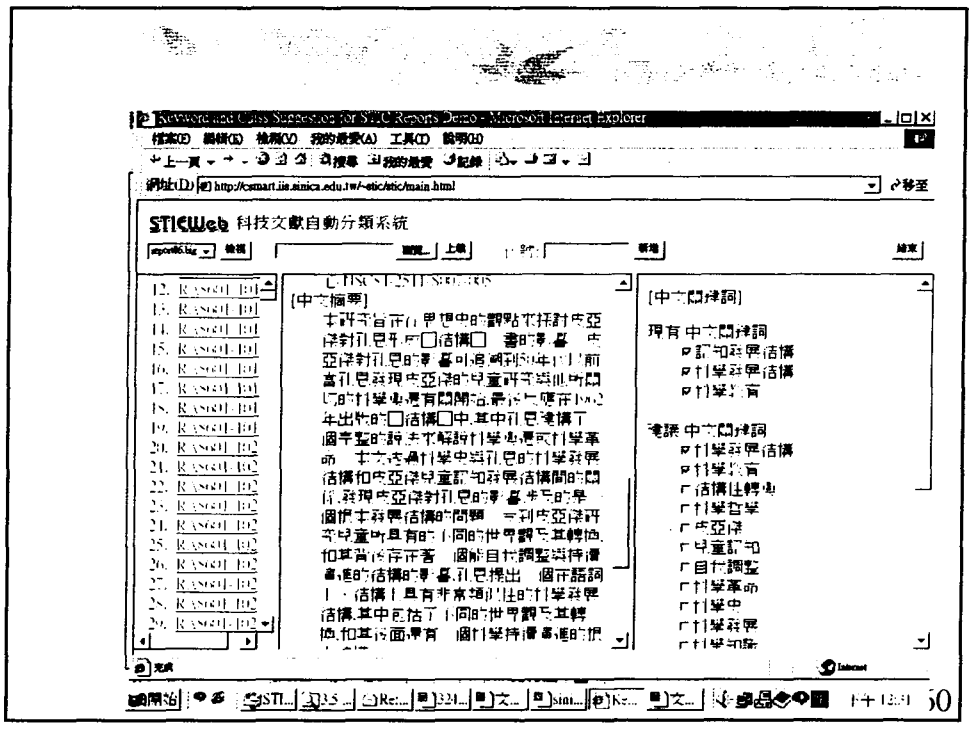
Cross-Language Web Search

- Allows users to query in one language and search documents written in another language.



58


Keyword and Class Suggestion for STIC Reports Demo - Microsoft Internet Explorer
STICWeb 科技文獻自動分類系統
[中文摘要]
[中文關鍵詞]
[英文關鍵詞]
59


Keyword and Class Suggestion for STIC Reports Demo - Microsoft Internet Explorer
STICWeb 科技文獻自動分類系統
[中文摘要]
[中文關鍵詞]
50

References

- **Web Mining**
- Kosala, R., & Blockheel, H. (2000). Web Mining Research: A Survey. *SIGKDD Explorations, 2(1),*1-15. PS PDF
  Srivastava,J. Cooley, R., Deshpande, M., & Tan, P.-N. (2000). Web usage mining:discovery and application of usage patterns from web data. *SIGKDD Explorations,1,* 12-23. PS
- J. Sirvastava & R. Cooley, Mining web data for e-commerce: concepts & applications, PKDD'01
- **Conferences & Workshops**
- KDD 2001, PKDD 2001, WebKDD 1999l, WebKDD 2000, WebKDD 2001
- **Web Content Mining**
- D. Mladenic et al., Text Mining: What if your data is made of words, PKDD'01
- M. Hearst, Untangling Text Data Mining, ACL'99.
- (Chang et al., 2001) (s.a.) Chapter 6 Handapparat
  Chakrabarti, S. (2000). Data mining for hypertext: A tutorial survey. *SIGKDD Explorations 1(2),* 1-11. PS PDF
- **Web Structure Mining**
- (Chang et al., 2001) (s.a.) Chapter 7.3 Handapparat
  (Chakrabarti, 2000) s.a.
- Page, L., Brin, S., Motwani, R.,& Winograd, T. (1998). *The PageRank Citation Ranking: Bringing Order to the Web.* PS
- **Web Usage Mining**
- (Srivastava et al., 2000) s.a.
  Spiliopoulou, M. (2000). Web usage mining for site evaluation: Making a site better fit its users. *Special Section of the Communications of ACM on "Personalization Technologies with DataMining", 43(8),* 127-134. Handapparat ACM Digital Library
- Cooley, R. 2000. *Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data.* University of Minnesotal. PS
- Borges, J.L. (2000).*A Data Mining Model to Capture User Web Navigation Patterns.* Department of Computer Science, University College London, London University. PS PDF
- For more references can refer at http://www.wiwi.hu-berlin.de/~berendt/lehre/2001w/wmi/literature.html

61

References (work in Web page classification)

- Text and Web page categorization
  - **S. Chakrabarti, B. Dorm, and P. Indyk. Enhanced hypertext categorization using hyperlinks. SIGMOD'98, pp. 307-318, 1998.**
  - **J. M. Pierre, Practical issues for automated categorization of Web sites, ECDL 2000 Workshop on the Semantic Web, 2000.**
  - **C.Y. Quek. Classification of World Wide Web Documents. Senior Honors Thesis, School of Computer Science, CMU, May 1997.**
  - **Y. Yang and X. Liu. A re-examination of text categorization methods, SIGIR'99, pp. 42-49, 1999.**
- Web page classification applications
  - **C. Chekuri, M.H. Goldwasser, P. Raghavan, and E. Upfal. Web search using automatic classification. WWW'97.**
  - **M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery. Learning to extract symbolic knowledge from the World Wide Web. AAAI'98, pp. 509-516, 1998.**
  - **M. Diligenti, F.M. Coetzee, S. Lawrence, C.L. Giles, and M. Gori, Focused crawling using context graphs, VLDB2000, pp. 527-534, 2000.**
- Link and context analysis
  - **G. Attardi, A. Gulli, and F. Sebastiani. Automatic web page categorization by link and context analysis. Proceedings of THAI'99, European Symposium on Telematics, Hypermedia and Artificial Intelligence, pp. 105-119, 1999.**
  - **S. Brin and L. Page. The anatomy of large-scale hypertextual web search engine, WWW'98.**
  - **J. Dean and M. R. Henzinger. Finding related pages in the world wide web. WWW'99, pp. 389-401, 1999.**
  - **J. Kleinberg. Authoritative sources in a hyperlinked environment. Proceedings of the 9th annual ACM SIAM Symposium on Discrete Algorithms, pp. 668-677, 1998.**

62

Event Ontology

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-
schema#>
 <rdfs:Class rdf:ID="Event"> </rdfs:Class>
 <rdfs:Class rdf:ID="Agent"> </rdfs:Class>
 .....
<rdf:Property rdf:ID="EventType">
 <rdfs:domain rdf:resource="#Event"></rdfs:domain>
</rdf:Property>
<rdf:Property rdf:ID= "IsPartOf">
  <rdfs:domain rdf:resource="#Agent" ></rdfs:domain>
  <rdfs:domain rdf:resource="#Action" ></rdfs:domain>
  .....
  <rdfs:range rdf:resource="#Event"></rdfs:range>
</rdf:Property>
 .....
</rdf:RDF>
```

Event Structure

- "荆軻　刺殺　秦王"
  Agent　Verb　Theme
- "他　是　秦王　的　兒子"
  Agent　Be-Verb　Theme　TSubject
- "秦王命李信攻打燕"
  - "秦王命李信"
  - "李信攻打燕"
  - "秦王命攻打燕"

References (work in Academia Sinica)

- 1. S. L. Chuang, L. F. Chien, "Automatic Subject Categorization of Query Terms for Web Information Retrieval", accepted by *Decision Support System*, 2002.
- 2. Lee-Feng Chien, et al., "Incremental Extraction of Domain-Specific Terms from Online Text Collections", *Recent Advances in Computational Terminology*, ed. By D. Bourigault et al., 2001.
- 3. Lee-Feng Chien, "PAT-Tree-Based Adaptive Keyphrase Extraction for Intelligent Chinese Information Retrieval" , special issue on "Information Retrieval with Asian Languages", *Information Processing and Management*, Elsevier Press, 1999.
- 4. W. H. Lu, L. F. Chien, H. J. Lee, " Mining Anchor Texts for Translation of Web Queries", accepted by *ACM Trans on Asian Language Information Processing*, 2002.
- 5. W. H. Lu, L. F. Chien, S. J. Lee, "Web Anchor Text Mining for Translation of Web Queries", *IEEE Conference on Data Mining*, Nov., San Jose, 2001.
- 6. C. K. Huang, L. F. Chien, Y. J. Oyang, "Interactive Web Multimedia Search Using Query-Session-Based Query Expansion", *The 2001 Pacific Conference on Multimedia* (PCM2001), Oct., Beijing.
- 7. C. K. Huang, Y. J. Oyang, L. F. Chien, "A Contextual Term Suggestion Mechanism for Interactive Search", *The First Web Intelligence Conference* (WI'2001), Japan.
- 8. Lee-Feng Chien. PAT-Tree-Based Keyword Extraction for Chinese Information Retrieval, The 1997 *ACM SIGIR Conference*, Philadelphia, USA, 50-58 (SIGIR'97).

63

Semantic Web

郭耀煌 教授

成功大學資工系

Semantic Web

Yau-Hwang Kuo
CREDIT Center
National Cheng Kung University

Contents

- Background
- Principles of Semantic Web
- Architecture of Semantic Web
- Language Stack of Semantic Web
- Ontology
- Applications
- Research Topics
- Conclusion

2

Background

1. growing complexity in web space
   * scale、device types、media type。
2. simplicity of HTTP and HTML has caused bottlenecks that hinder searching, extracting, maintaining, and generating information。
3. readable to human → machine understandable。
4. knowledgeable usage of webs。
5. efficiency in handling web data。

3

Background (Cont.)

6. needs of service automation：browsing by users to retrieve information → automatically cooperating by webs to provide services。

So, we need the third generation webs。
(hand written HTML pages → machine generated HTML pages → semantic web) 。

4

Basic Principles

1. meaningful expression by structured data model and knowledge representation schemes ◦
2. machine readable and self-describing information and knowledge ◦
3. standardizing knowledge sharing and reusable on web ◦
4. knowledge acquisition and evolution by fast web space growing ◦
5. ontology as domain theory to maintain semantic consistency ◦
6. agent support ◦
7. interoperability across different platforms and devices ◦

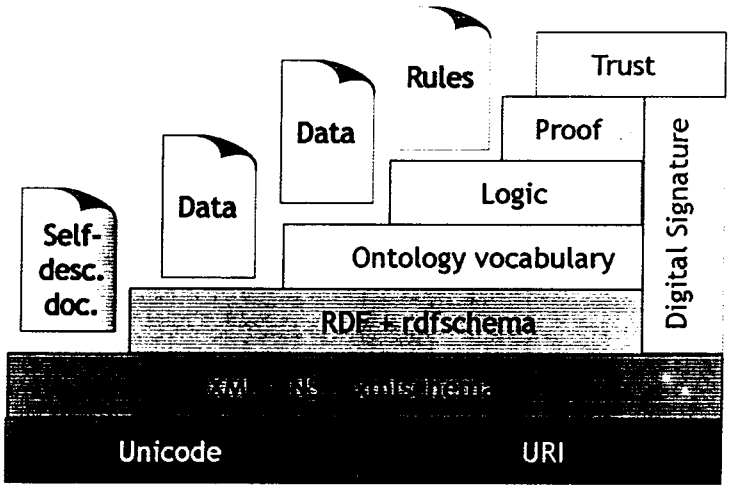
5

Layers of Semantic Web

- Unicode + URI (foundation) layer
- XML (syntactic interoperability) layer
- RDF + Schema (data interoperability) layer
- Ontology (data inter-conversion) layer
- Logic (interoperability) layer

6

Architecture of Semantic Web



7

Basic Assertion Model

- minimalist design: based on a common model of great generality →RDF.
- a very simple model that we can map the data in any new format.
- just contains the concepts of assertion and quotation (representation of metadata).
- needs layering many things on the top.
- Without great power but easy to combine information.

8

Schema Layer

- Class, property.
- Subproperty, subclass.
- Domain and range.
- Comments and labels.

9

Ontology Layer

- More metainformation, such as transitive property, unique, unambiguous, cardinality, etc.
- Not turing complete.
- Huge extra usage for extra functionality.

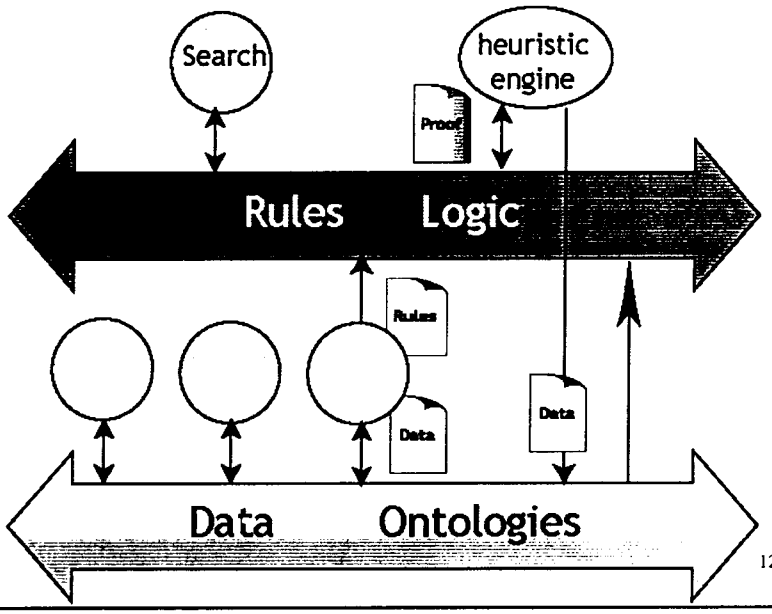
10

Logic layer

- Universal language for monotonic logic.
- No one standard engine – inference capabilities differ.
- Any system can validate proof.
- Logic: state any logic principle and permit the computer to reason using these principles.
- Proof: follow logic to prove things.
- Trust: digital signatures and web of trust.

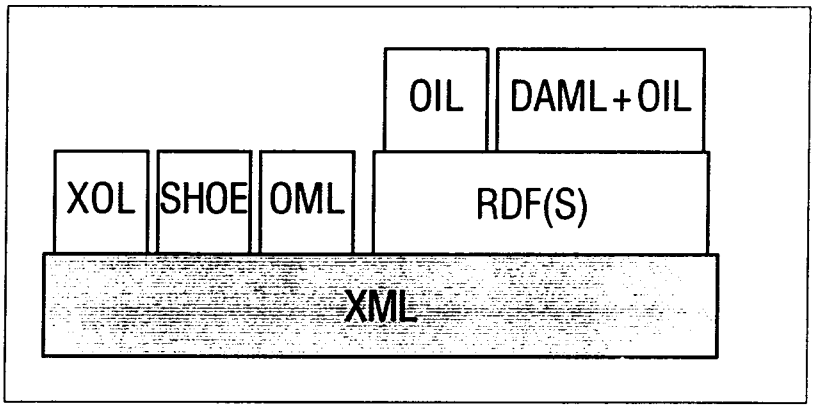
11

Semantic Web bus



12

Language stack in Semantic Web



13

Uniform Resource Identifier

- Identify items on the web and consider it as a resource.
- The foundation of web: while nearly every other part of the web can be replaced, the URI cannot; it holds the rest of the web together.
- URL: one form of URI; both identifies and locates.
- A "mid:" URI identifies an email message, but it isn't able to locate a copy of the message for you.

14

Uniform Resource Identifier (cont.)

- URIs are decentralized: while some URI schemes (http:) depend on centralized systems (such as DNS), other schemes (such as freenet:) are completely decentralized.
- You don't need anyone's permission to create a URI →this flexibility makes URIs powerful.
- Inevitably end up with multiple URIs representing the same thing.

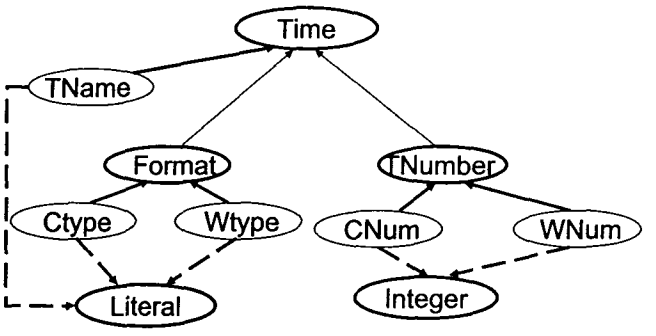
15

Uniform Resource Identifier (cont.)

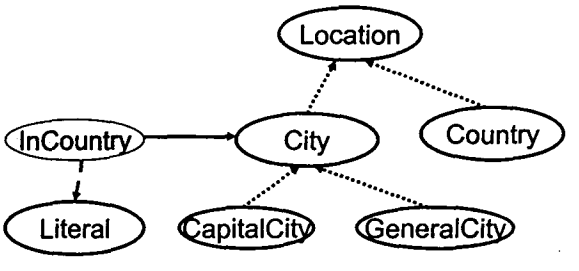
- A common practice for creating URIs is to begin with a web page: this page describes the object to be identified and explains that the URL of the page is the URI for that object.
- URI may name a resource not accessible over the Internet. While URL is a type of URI that does provide a way to get information about a resource, or perhaps to retrieve the resource itself.

16

Time ontology (Schema)



Location ontology (Schema)



RDF and RDF Schema

- Developed by W3C for describing Web resources, allows the specification of the semantics of data based on XML in a standardized, interoperable manner.
- It also provides mechanisms to explicitly represent services, processes, and business models, while allowing recognition of nonexplicit information.

17

RDF and RDF Schema (cont.)

- Equivalent to the semantic network formalism.
- Consists of three object types:

  1. resources: described by RDF expressions and always named by URIs.

  2. properties: define specific aspects, characteristics, attributes, or relations to describe a resources.

  3. statements: assign a value for a property in a specific resource (this value might be another RDF statement).

18

RDF and RDF Schema (cont.)

- Basically, RDF is based on O-A-V representation scheme.
- RDF does not provide mechanisms for defining the relationships between properties (attributes) and resources.
- RDFS offers primitives for defining knowledge models that are closer to frame-based approaches.
- Protégé, Mozilla, Amaya, etc. adopt RDF(s).

19

XOL: Ontology Exchange Language

- Designed for the exchange of ontology definition among a heterogeneous set of software systems in their domain.
- Designed by US bioinformatics community.
- Based on Ontolingua and OML.\
- XOL files use XML syntax, so that we can use an XML editor to author XOL files.

20

SHOE: Simple HTML Ontology Extension

- Develop at the University of Maryland and used to develop OML, was created as an extension of HTML, incorporating machine-readable semantic knowledge in HTML documents or other web documents.
- Recently, Maryland has adapted the SHOE syntax to XML.

21

SHOE: Simple HTML Ontology Extension

- SHOE makes it possible for agents to gather meaningful information about Web pages and documents, improving search mechanisms, and knowledge gathering.
- This process consists of 3 phases:

  1. define an ontology,

  2. annote HTML pages with ontological information,

  3. retrieve web information by an agent.

22

OML: Ontology Markup Language

- Developed at the Univ. of Washington.
- Partially based on SHOE.
- 4 levels of OML:

  1. OML core: related to logical aspects of the language and is included by the rest of the layers.
  2. Simple OML: maps directly to RDF(S).

23

OML: Ontology Markup Language

3. abbreviated OML: includes conceptual graphs features.
4. standard OML: the most expressive version of OML.

24

OIL: Ontology Interchange Language

- Developed in the OntoKnowledge project.
- Permit semantic interoperability between Web resources.
- Syntax and semantics are based on existing proposals (OKBC, XOL, and RDF(S)).
- Provide modeling primitives commonly used in frame-based approaches to ontological eng. and formal semantics and reasoning support found in description logic approaches.

25

OIL: Ontology Interchange Language

- Layers of OIL:

1. Core OIL: groups the OIL primitives that have a direct mapping to RDF(S) primitives.
2. Standard OIL: the complete OIL model; use more primitives that the ones defined in RDF.
3. Instance OIL: adds instances of concepts and roles to the previous model.
4. Heavy OIL: future extension of OIL.

26

DARPA Agent Markup Language + OIL

- Developed by a joint committee from the US and European Union (IST).
- Based on DAML project for allowing semantic interoperability in XML.
- Protege2000, OntoEdit and WebODE are tools that can author DAML+OIL ontologies.

27

Comparisons Criteria

1. Concepts
2. Taxonomies
3. Relations
4. Functions
5. Axioms
6. Instances

28

Comparison in Concepts: Table 1

Concepts	XOL	SHOE	OML	RDF(S)	OIL	DAML+OIL
General Issues						
Partitions	-	-	+	-	+	+
Documentation	+	+	+	+	+	+
Attributes						
Instance attributes	+	+	+	+	+	+
Class attributes	+	-	+	-	+	+
Local scope	+	+	+	+	+	+
Global scope	+	-	+	+	+	+
Facets						
Default value	+	-	-	-	-	-
Type constraints	+	+	+	+	+	+
Cardinality constraints	+	-	-	-	+	+
Documentation	+	+	+	+	+	+

29

Comparison in Taxonomies: Table 2

Taxonomies	XOL	SHOE	OML	RDF(S)	OIL	DAML+OIL
Subclass of	+	+	+	+	+	+
Exhaustive decompositions	-	-	+/-	-	+	+
Disjoint decompositions	-	-	+	-	+	+
Not subclass of	-	-	-	-	+	+

30

Comparison in Relations and Functions: Table 3

Relations and functions	XOL	SHOE	OML	RDF(S)	OIL	DAML+OIL
n-*ary* relations/functions	+/-	+	+	+/-	+/-	+/-
Type constraints	+	+	+	+	+	+
Integrity constraints	-	-	+	-	-	-
Operational definitions	-	-	-	-	-	-

31

Comparison in Axioms: Table 4

Axioms	XOL	SHOE	OML	RDF(S)	OIL	DAML+OIL
First-order logic	-	+/-	+	-	+/-	+/-
Second-order logic	-	-	-	-	-	-
Independent axioms	-	-	-	-	-	-
Embedded axioms	-	-	+	-	-	-

32

Comparison in Instances: Table 5

Instances	XOL	SHOE	OML	RDF(S)	OIL	DAML+OIL
Instances of concepts	+	+	+	+	+	+
Facts	+	+	+	+	+	+
Claims	-	+	-	+/-	+/-	+/-

33

Ontology: A Revolution for Information Access and Integration

- An ontology is a formal, explicit specification of a shared conceptualization
  - *Conceptualization* refers to an abstract model of some phenomenon in the world that identifies that phenomenon's relevant concepts.
  - *Explicit* means that the type of concepts used and the constraints on their use are explicitly defined.
  - *Formal* means that the ontology should be machine understandable
- The main application areas of ontology technology
  - Knowledge management
  - Web commerce
  - Electronic business

34

What is Ontology?

- Ontology – explicit formal specifications of the terms in the domain and relations among them.
- An ontology contains a hierarchy of concepts within a domain and describes each concept's property through an attribute-value mechanism.
- Relations between concepts describe additional logical sentence.

35

Characteristics of Ontology

1. Formal Semantics
2. Consensus of terms
3. Machine readable and processible
4. Model of real world
5. Domain specific

36

Time and Location schema

④“西元前　227　年”
Wtype　WNum
– “在　長平之戰　期間“
TName
– “　秦　都城　咸陽”
Country/InCountry　CapitalCity

A Simple Sentence

– a sentence with only one verb.
– only deal with transitive verb and be-verb
– A grammar of a tuple (Agent, Verb, Theme) is similar to (Subject, VP, NP)

(Chinese),秦將軍李信攻打燕於西元前226年
(English),The general of Chin Dyansty,Li-Ching,
attacked Yen Country in 226 B.C.

Reasons to Develop Ontologies

- To share common understanding of the structure of information among people or software agents.
- To enable reuse of domain knowledge.
- To make domain assumptions explicit.
- To separate domain knowledge from the operational knowledge.
- To analyze domain knowledge.

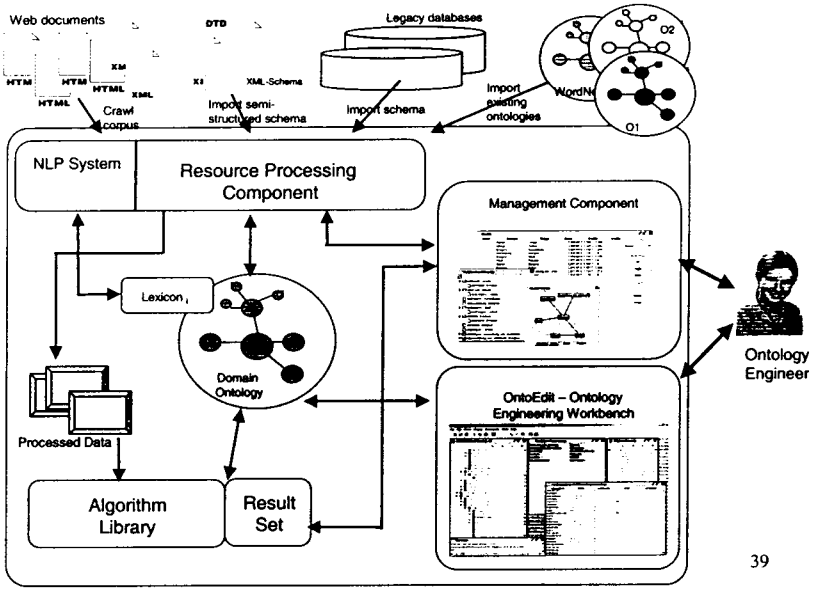
37

Process of Developing an Ontology

- Developing an ontology includes:
  - Determine the domain and scope of the ontology.
  - Consider reusing existing ontologies.
  - Enumerate important terms in the ontology.
  - Define classes in the ontology and arrange the classes in a taxonomic (subclass-superclass) hierarchy.
  - Define attribute and describe allowed values for these attribute.
  - Fill in the values for attribute for instance.

38

Ontology-learning Architecture for Semantic Web



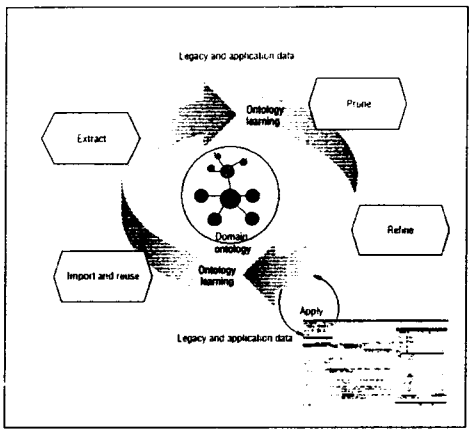
39

Ontology-learning Architecture (Cont.)

- Flexible handling of large amounts of source is done within the *Management component*
- Importing and reusing existing structures or ontology using *Resource processing component*
- Natural Language texts have to be processed using natural language processing techniques: *NLP System (SMES – Saarbrücken Message Extraction System)*
- Algorithms should discover conceptual structures => *Algorithm Library* with an evaluation module is required
- Humans have to judge about the interestingness and the final position in the ontology of the discovered conceptual structures => *Ontology Engineering Environment OntoEdit*

40

Ontology Learning Process



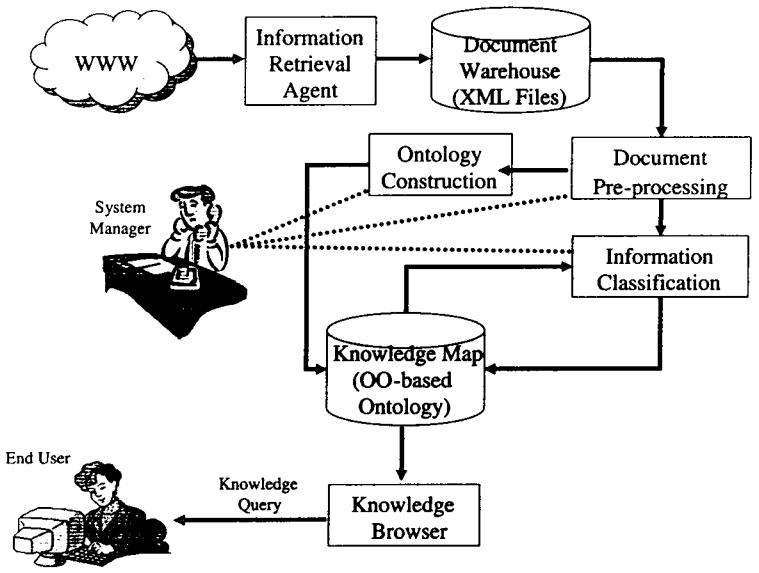
41

Automatic Construction of OO Ontology

1. Use object-oriented data model to represent ontologies ◦
2. Follow object-oriented analysis procedure to build ontologies
3. Apply natural language processing technology to extract key terms from documents ◦
4. Apply fuzzy clustering techniques to find concepts ◦
5. Apply data mining techniques to associate concepts ◦
6. Apply NLP techniques to extract attributes and operations of concepts ◦
7. Apply OO approach to build generalization and aggregation hierarchy ◦

42

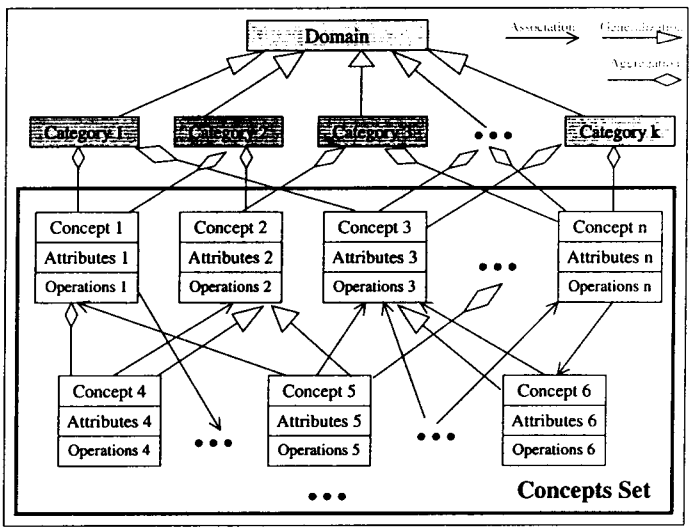
系統架構圖



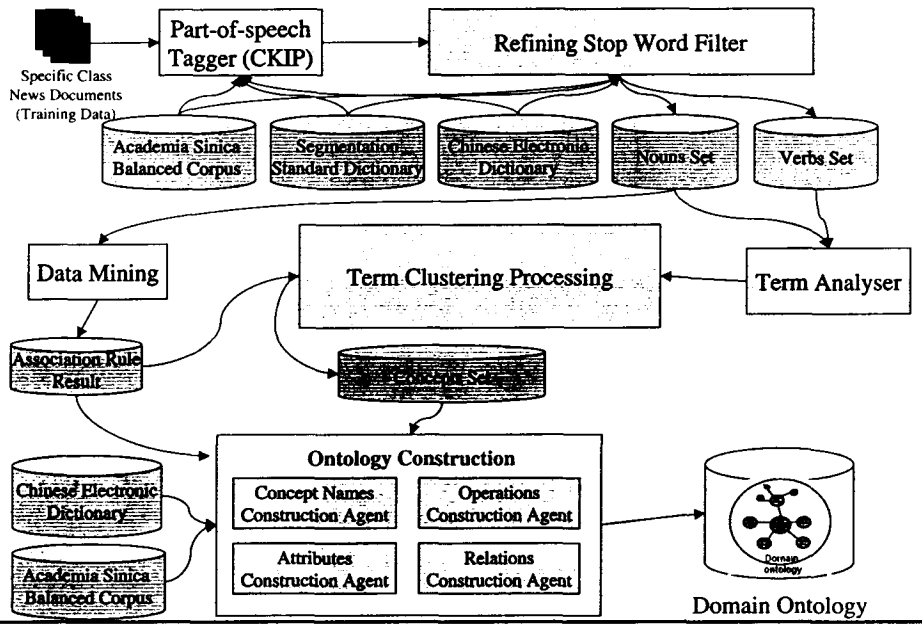
43

Structure of Object-Oriented Ontology

- Domain
- Category
- Concepts
  - Concept names
  - Attributes
  - Operations
- Relations
  - Association
  - Generalization
  - Aggregation



Ontology 自動建構流程圖



Domain Ontology

Part-of-speech Tagging (CKIP)

	Part-of-Speech tag	Meaning
1.	VA	Intransitive verb
2.	VB	Single verb class
3.	VC	Single verb
4.	VD	Two words verb
5.	VE	Verb sentence
6.	VF	Name verb
7.	VG	Classification verb
8.	VH	Status Intransitive verb
9.	VI	Status single verbs class
10.	VJ	Status single verbs
11.	VK	Status single verbs
12.	VL	Status name verb
13.	A	Adjective
14.	Na	Common noun
15.	Nb	Proper noun
16.	Nc	Location noun
17.	Nd	Time noun
18.	Ne	Stable noun
19.	Nf	Quantity noun
20.	Ng	Direction noun
21.	Nh	Pronoun
22.	Da	Adverb, Quantity
23.	Dba	Law Adverb

24.	Dbb,Dbc	Evaluation Adverb
25.	Dc	Negate Adverb
26.	Dd	Time Adverb
27.	Df	Degree Adverb
28.	Dg	Location Adverb
29.	Dh	Method Adverb
30.	Di	Symbol Adverb
31.	Dj	Interrogative Adverb
32.	Dk	Sentence Adverb
33.	P	Preposition
34.	Ca	Side by side conjunction
35.	Cb	Correlation conjunction
36.	T	Particle
37.	I	Interjection

The Part-of-Speech tags and their corresponding meanings of CKIP

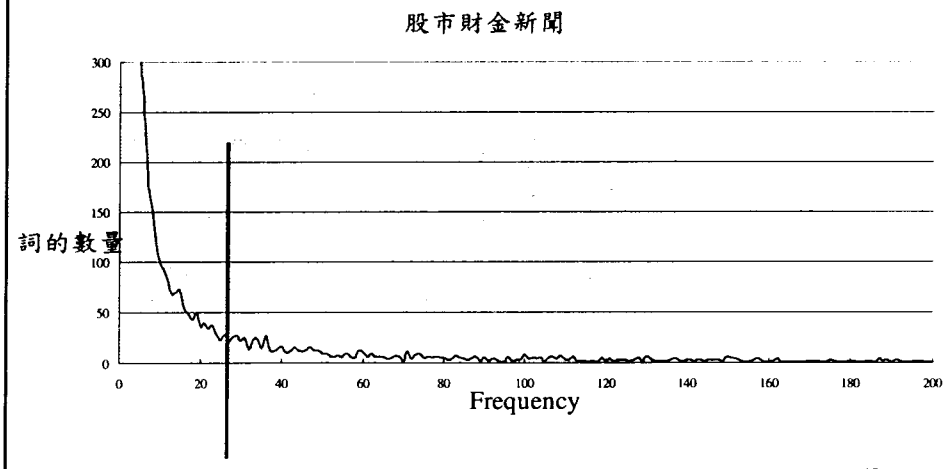
46

Stop Word Filter

- 參考中文詞庫，作更精確的詞性定義
  - Na: Naa Nab Nac Nad …
  - Nb: Nba Nbc …
  - Nc: Nca Ncb Ncc …
  - …
- 保留詞之詞性
  - 普通名詞(Na)，專有名詞(Nb)，地方名詞(Nc)
- 過濾詞之詞性
  - 時間名詞(Hd)，定詞(Ne)，量詞(Nf)，方位詞(Ng)，代名詞(Nh)

47

詞的頻率分析



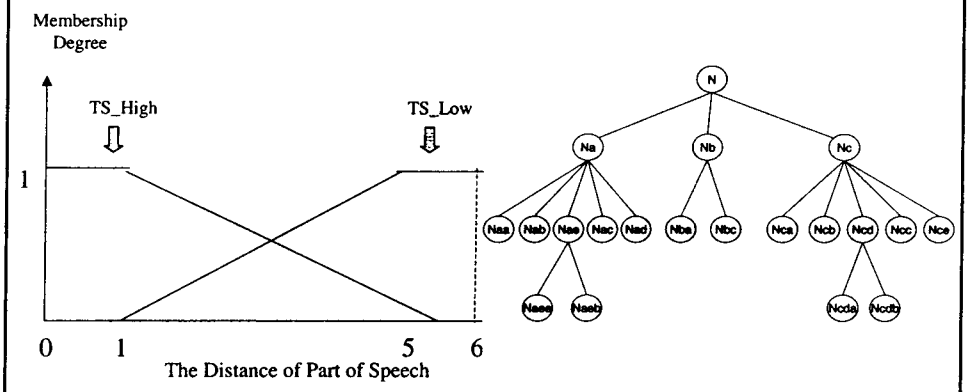
48

Evaluation of Conceptual Resonance of Terms

- Conceptual resonance means the degree of denoting the same concept between two different terms. So, two terms will have a high opportunity to be represented with the same concept.
  - Resonance in Part of Speech
  - Resonance in Term Vocabulary
  - Resonance in Term Association
  - Resonance in Common Term Association
  - Aggregate Term Resonance

49

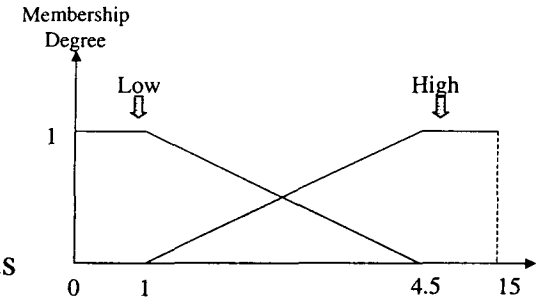
Resonance in Part of Speech



50

Resonance in Term Vocabulary

- Evaluation in Common Words
  - 民進黨，民進黨團，民主進步黨
  - 海峽兩岸、兩岸、兩岸政策
- Evaluation in Common Starting or Ending Words
  - 台北縣，台南縣
  - 警員，警察，警方

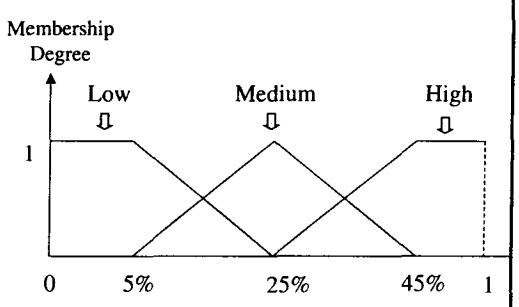


The Similarity of Term Vocabulary

51

Resonance in Term Association

- Clustering just based on term knowledge
  - {總統，總統府}，{陳水扁}
  - {台北市政府，台北市長}，{馬英九}
- Consider in Term Association
  - Confidence(總統->陳水扁) {總統，總統府，陳水扁}
  - Confidence (台北市長->馬英九) {台北市政府，台北市長，馬英九}

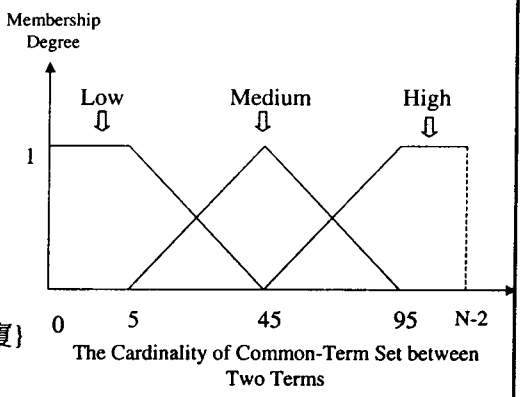


The Confidence Value between Two Terms

52

Resonance in Common Term Association

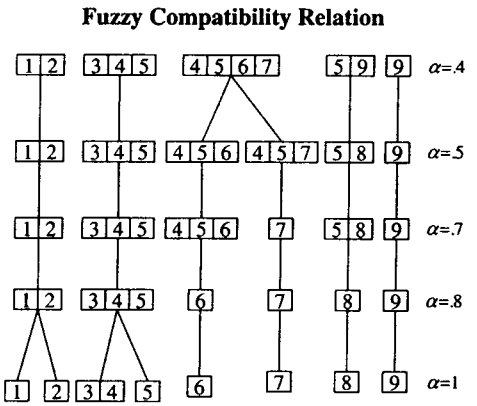
- Clustering just based on term knowledge
  - {警方，美方} or {美國，美方} *contradictive ?*
- Common term association analysis
  - Threshold of Confidence
  - 美國 -> {白宮，布希，紐約}
  - 美方 -> {白宮，布希，五角大廈}
  - 警方 -> {警員，刑事組，分局}
- {警方}，{美方，美國}



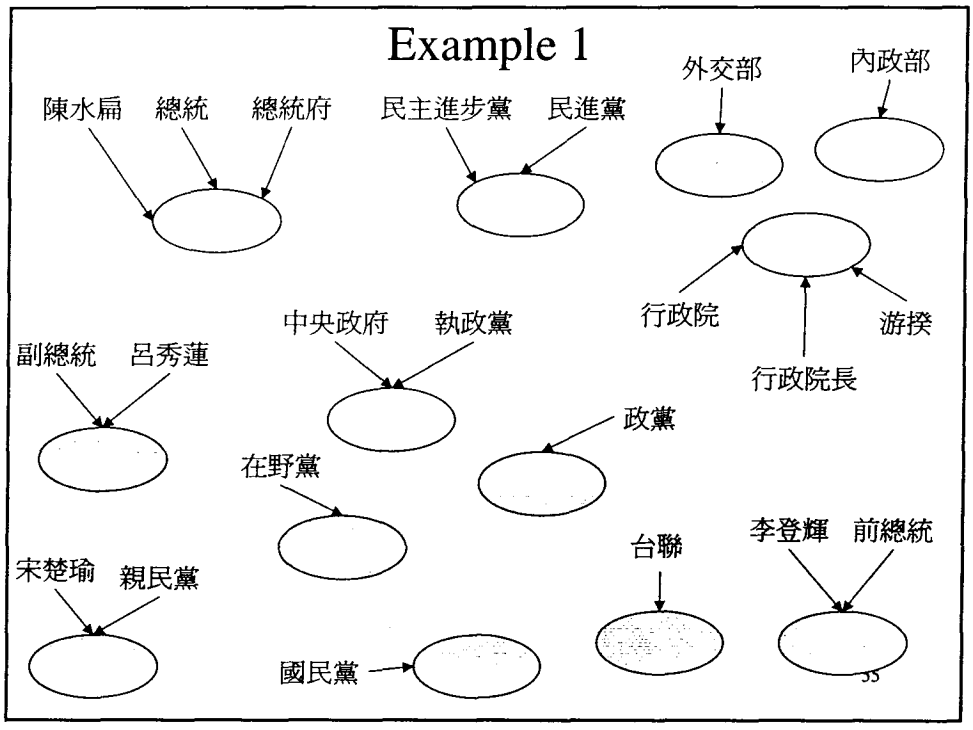
53

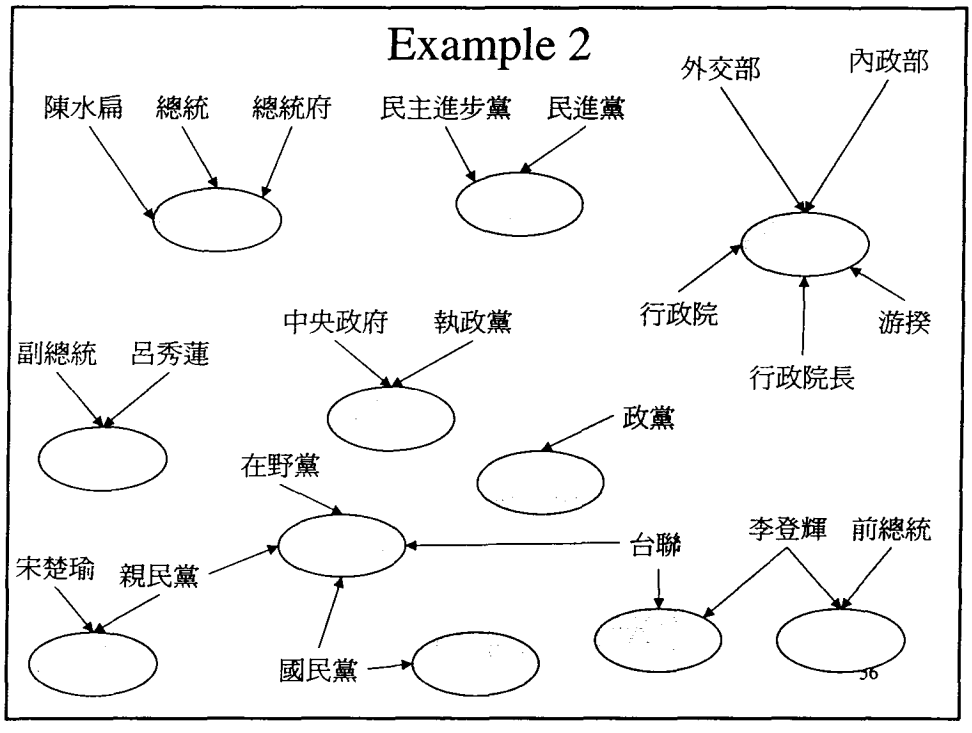
Concept Clustering based on Fuzzy Compatibility Relation Approach

- 可將Conceptual Resonance of Terms 視爲Fuzzy Compatibility Relation
  - Reflexive
  - Symmetric
- Decide Compatibility Membership Degree α



54


Example 1
陳水扁
總統
總統府
民主進步黨
民進黨
外交部
內政部
行政院
游揆
行政院長
副總統
呂秀蓮
中央政府
執政黨
政黨
在野黨
台聯
李登輝
前總統
宋楚瑜
親民黨
國民黨
55


Example 2
陳水扁
總統
總統府
民主進步黨
民進黨
外交部
內政部
行政院
游揆
行政院長
副總統
呂秀蓮
中央政府
執政黨
政黨
在野黨
台聯
李登輝
前總統
宋楚瑜
親民黨
國民黨
56

A Simple Sentence in RDF

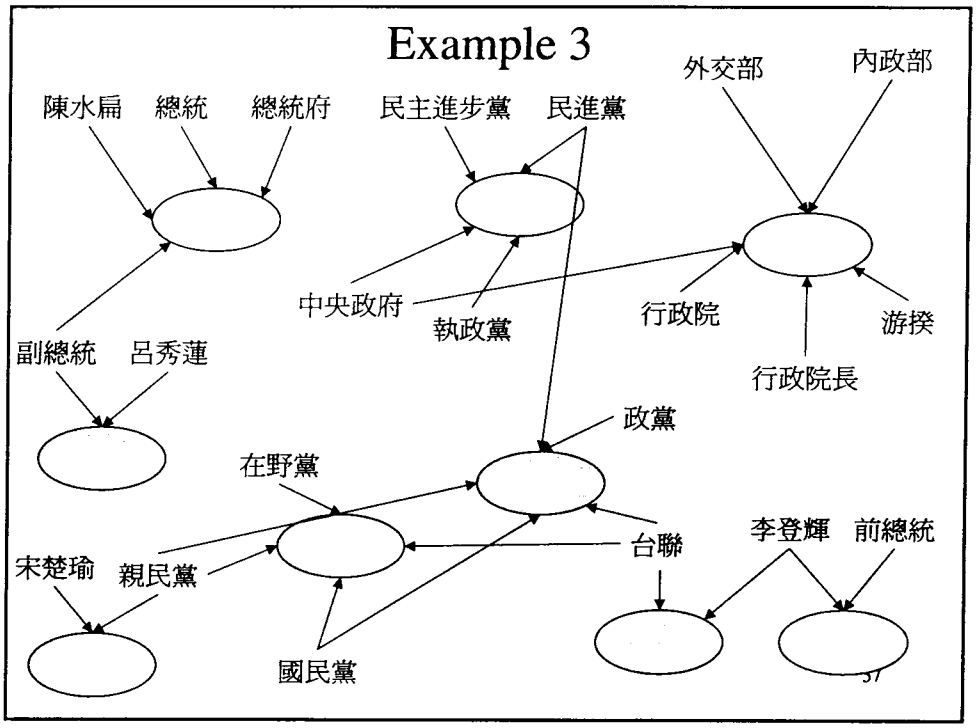
```
.....
xmlns:s="http://aidl.cs.nthu.edu.tw/idlp/event_ontology#" >
.....
<s:Agent rdf:ID="李信">
  <s:a_IsPerson>是</s:a_IsPerson>
  <s:a_Nationality>秦</s:a_Nationality>
  <s:a_Identity>將軍</s:a_Identity>
</s:Agent>
<s:Action rdf:ID="Action01">
  <s:Verb>攻打</s:Verb>
</s:Action>
.....
<s:Time rdf:ID="西元前226年">
  .....
    <s:Wtype>西元前</s:Wtype>
    .....
    <s:WNum>226</s:Wnum>
  .....
</s:Time>
.....
</rdf:RDF>
```

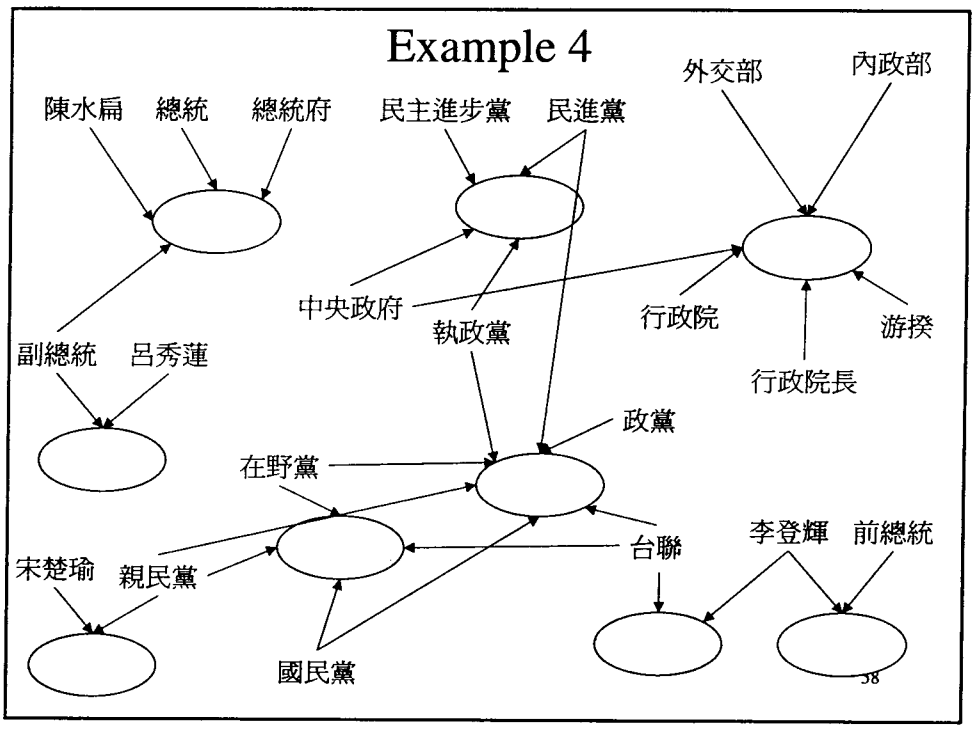
Linguistic Analysis of Sentences

Original:
秦始皇是秦襄王之子，於西元前二二一年滅了其以後，建立了一個中央集權的秦國。

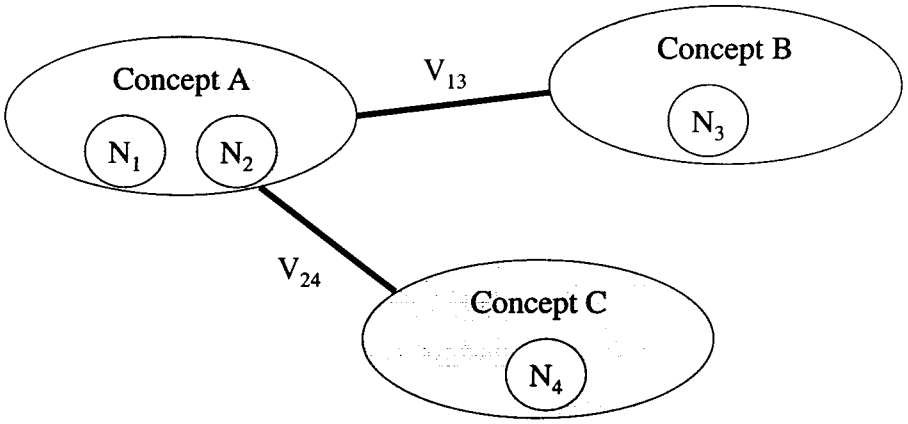
Result:
秦始皇是秦襄王之子， 西元前二二一年滅齊， 建立秦國。

"秦始皇" is the subject of "是", "滅", and "建立".


Example 3
陳水扁
總統
總統府
民主進步黨
民進黨
外交部
內政部
中央政府
執政黨
行政院
游揆
行政院長
副總統
呂秀蓮
政黨
在野黨
台聯
李登輝
前總統
宋楚瑜
親民黨
國民黨
57


Example 4
陳水扁
總統
總統府
民主進步黨
民進黨
外交部
內政部
中央政府
執政黨
行政院
游揆
行政院長
副總統
呂秀蓮
政黨
在野黨
台聯
李登輝
前總統
宋楚瑜
親民黨
國民黨
58

Associate Concepts by Data Mining based Approach

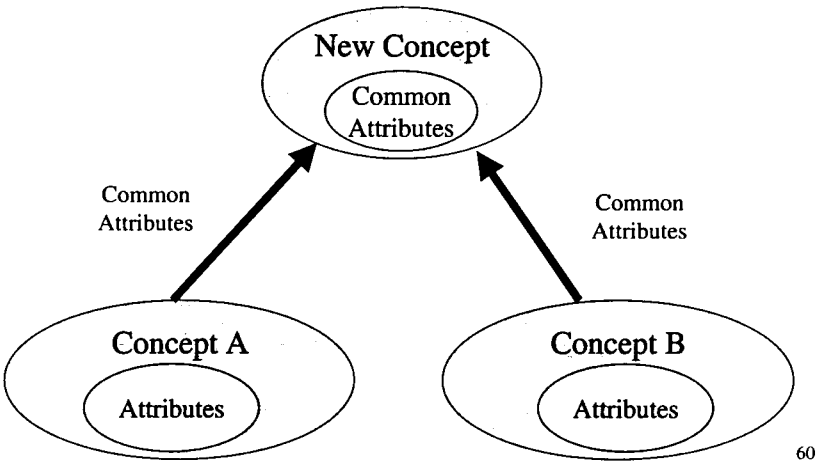


Support (N_1, N_3) → Confidence (V_{13})

Support (N_2, N_4)→ Confidence (V_{24})

59

Building Generalization Hierarchy and Aggregation Hierarchy



60

Research Topics

1. Languages ◦
2. Data Models ◦
3. Tools (authoring、visualization、converting) ◦
4. Automatic construction、learning and adaption of ontologies ◦
5. Consistency maintenance of ontologies ◦
6. Cooperation and negotiation of multiple agents ◦
7. Reasoning over webs ◦
8. Decentralized and heterogeneous web knowledge processing ◦
9. Service discovery and composition ◦
10. Logic and proof ◦

61

Applications

1. Knowledge Management
2. Personalized Information Brokers
3. Electronic business
4. One-to-One Web commerce
5. Personalized Activity Scheduling
6. Document Management
7. Personalized Finance Support
8. Cyber Circles
9. Location-based service discovery and Composition

62

Web Services Technology and Applications

楊仁達 博士

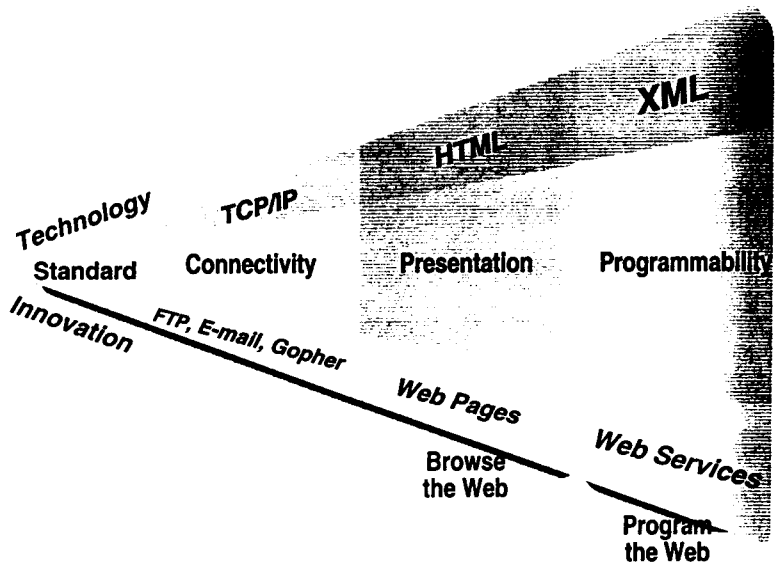
資策會

Web Services Technology and Application

楊仁達

財團法人資訊工業策進會
Act 電子商務技術實驗室
Advanced e-Commerce Technology Laboratory

Next Wave: XML Web Services



Agenda

- Part I: Business prospect
- Part II: Technical prospect

Business prospect

Part I

Business prospect

- Web Services Overview
  - What are Web Services?
  - Why Web Services?
  - How will Web Services reshape the business?
- Web Services and IT application
  - EAI
  - B2Bi
  - Content integration
  - Service integration
- Web Services Business Model
  - Hyper vs. reality
- Case Study
  - Microsoft .Net
  - IBM
  - Others

Definitions of Web Services

- Microsoft
  - A Web service is programmable application logic, accessible using standard Internet protocols
  - A standards based messaging architecture to loosely couple applications in a decentralized environment
- IBM
  - Web services are self-contained, modular applications that can be described, published, located, and invoked over a network, generally, the World Wide Web
- Gartner Group
  - A software component that represents a business function (or a business service) and can be accessed by another application (a client, a server or another Web service) over public networks using generally available ubiquitous protocols and transports (i.e. SOAP over HTTP)."

What is a Web Service?

- A Capability *(Web **Service**)*
- Accessed via Internet Protocols *(**Web** Service)*
- Using Open Standards
- Can be dynamically found and included in application
- Provides programmable access
- Enabling information to be queried, updated in real-time
- Improving efficiency, integrity, responsiveness and accuracy



B2C

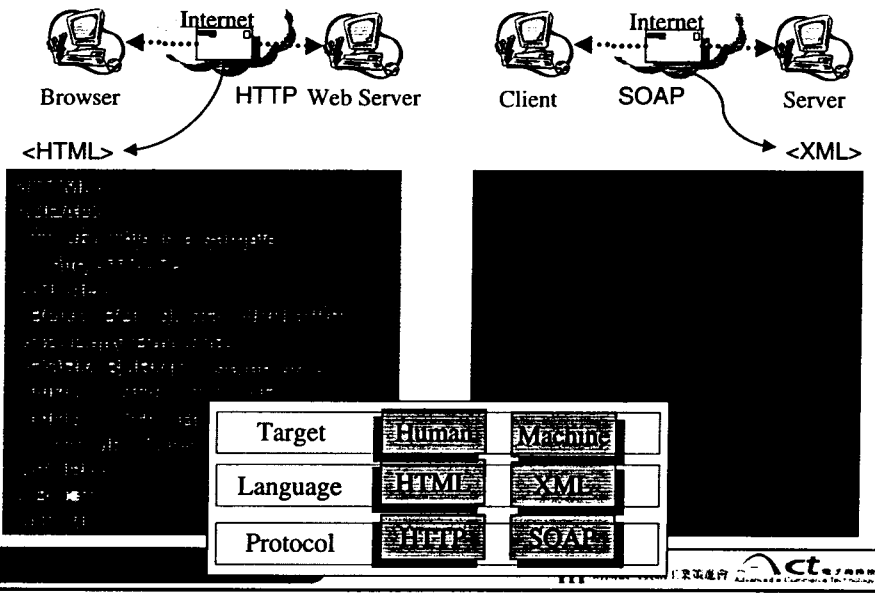


B2B



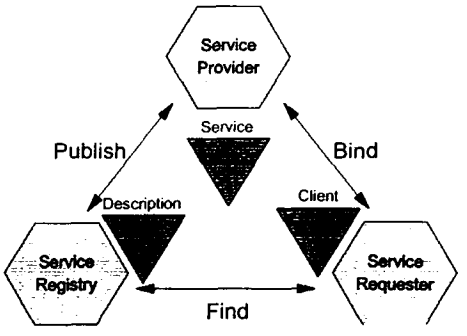
Source: CBDi

Information Exchange by standards

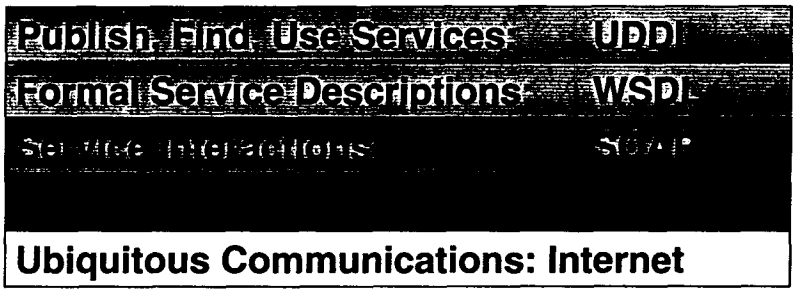


The Elements of a Web Service

- Key Players
  - **The Service Provider**
  - **The Service Requester**
  - **The Service Registry**
- Key Functions
  - **Publish**
  - **Find**
  - **Bound**
- All through open standard
  - Loosely couple and dynamic binding
  - Cross Platforms / Cross devices



Web Services Basic Stack



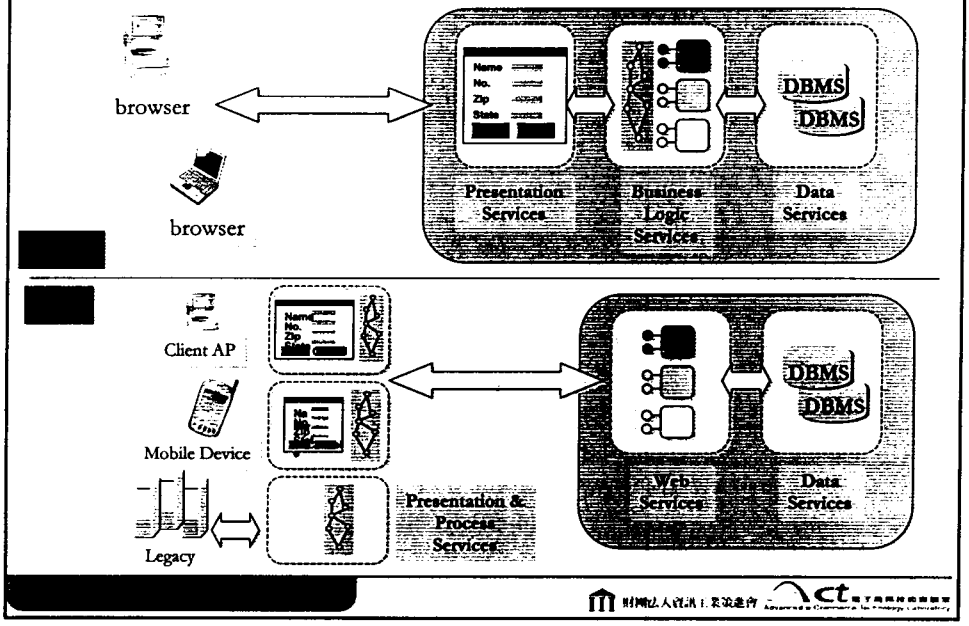
UDDI : Universal Description Discovery and Integration

WSDL: Web Service Description Language

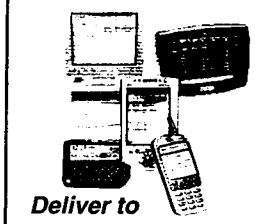
SOAP : Simple Object Access Protocol

Simple, Open, Broad Industry Support

Why Web Services?



Examples – Retailer to Customer



Deliver to applications running on variety of devises

List Product Catalog

Notify Special Offers

Order with 'e-code'

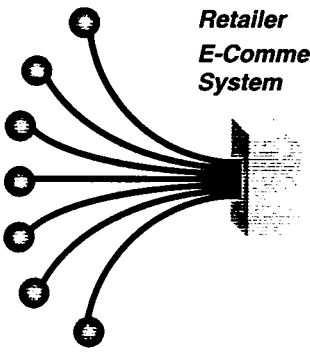
Query Order Status

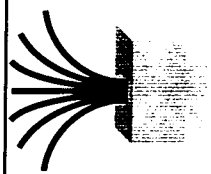
Place bid on Auction

Notify Delivery

Notify Item Back in Stock

Retailer E-Commerce System





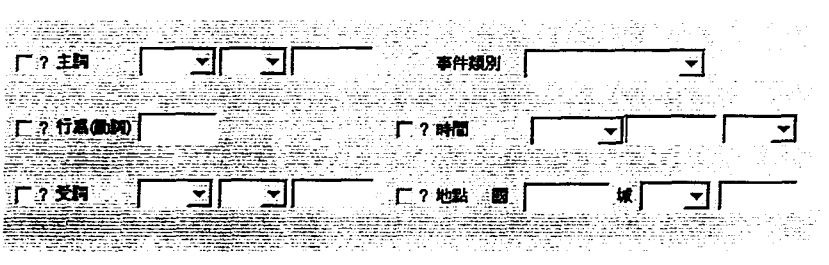
Deliver to application of another intermediaries

Source: CBDi

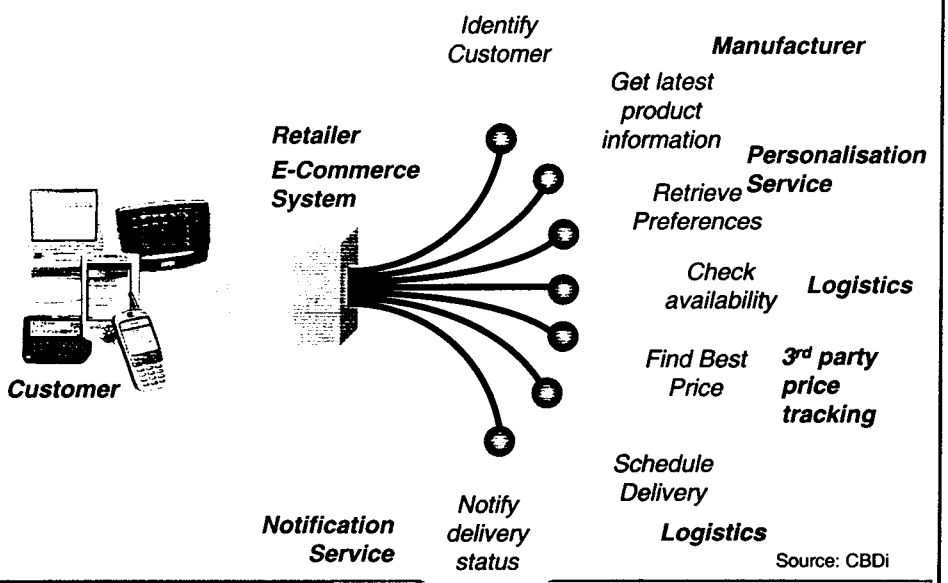
Query representation

- We use some selection functions for users to fulfill what might related to their queries by choosing the suitable items.
- Understanding the requirements of users becomes more consistent and less effort.

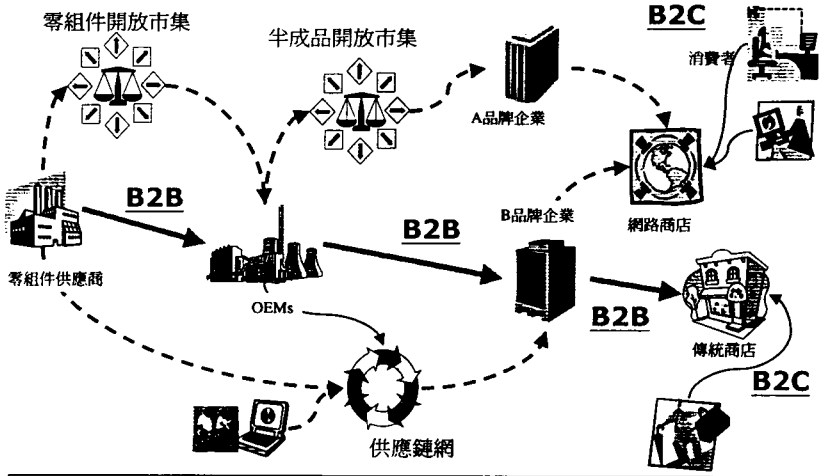
Query Template on Interface



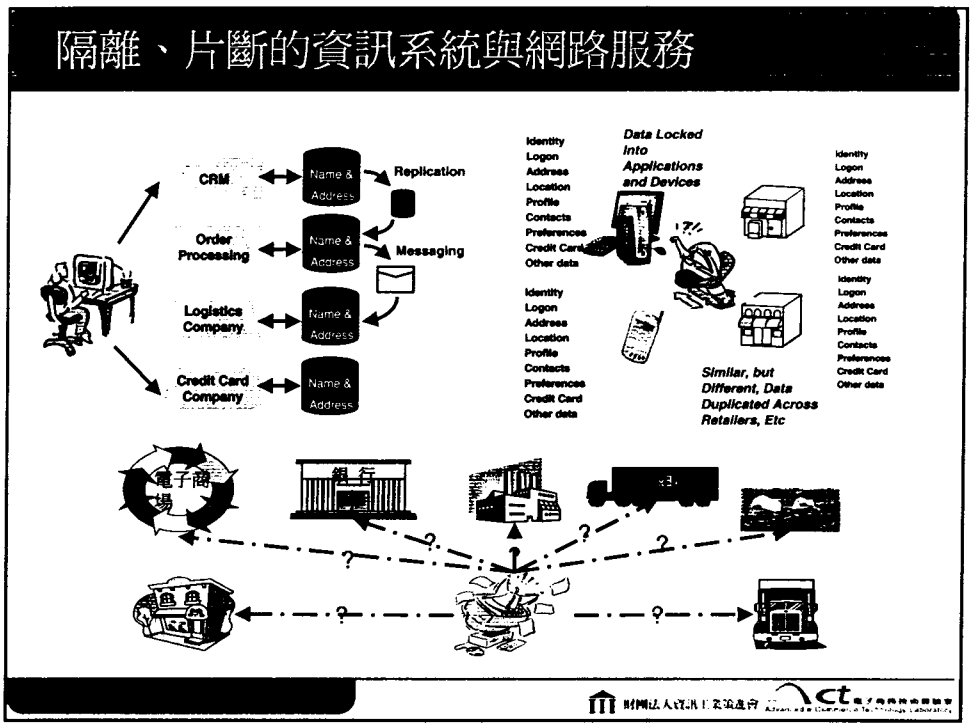
Examples – Retailer from Suppliers and Partners

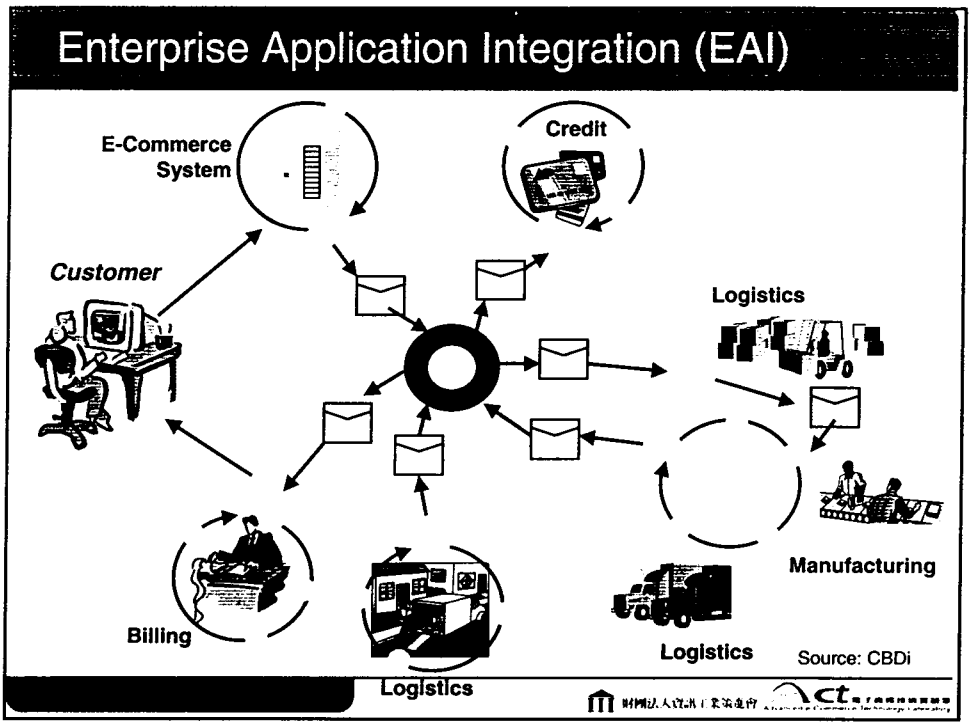


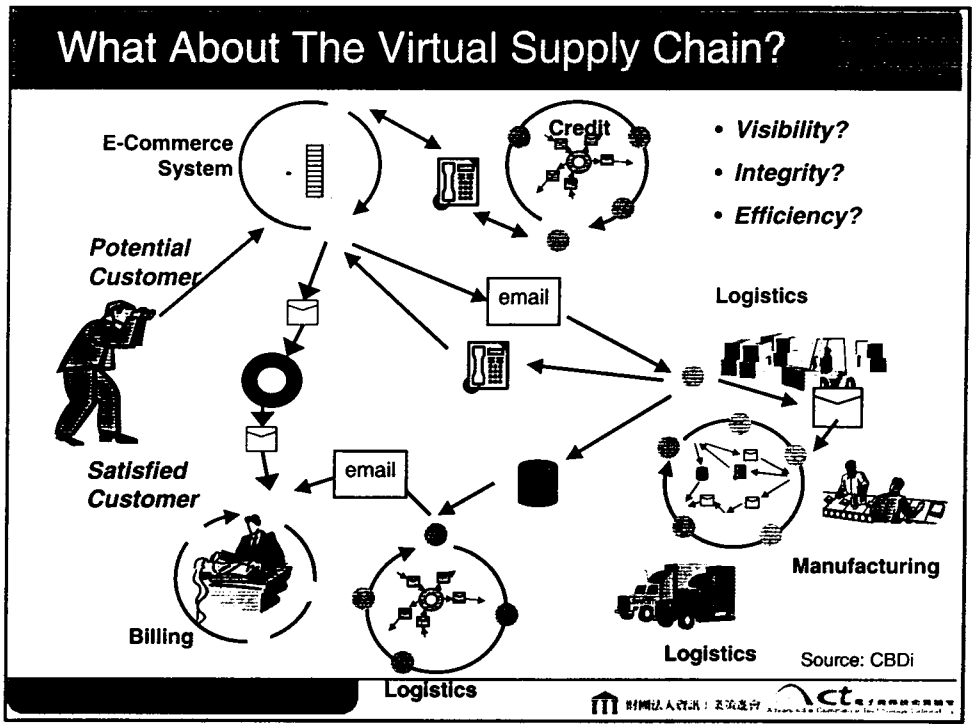
整合價值鏈

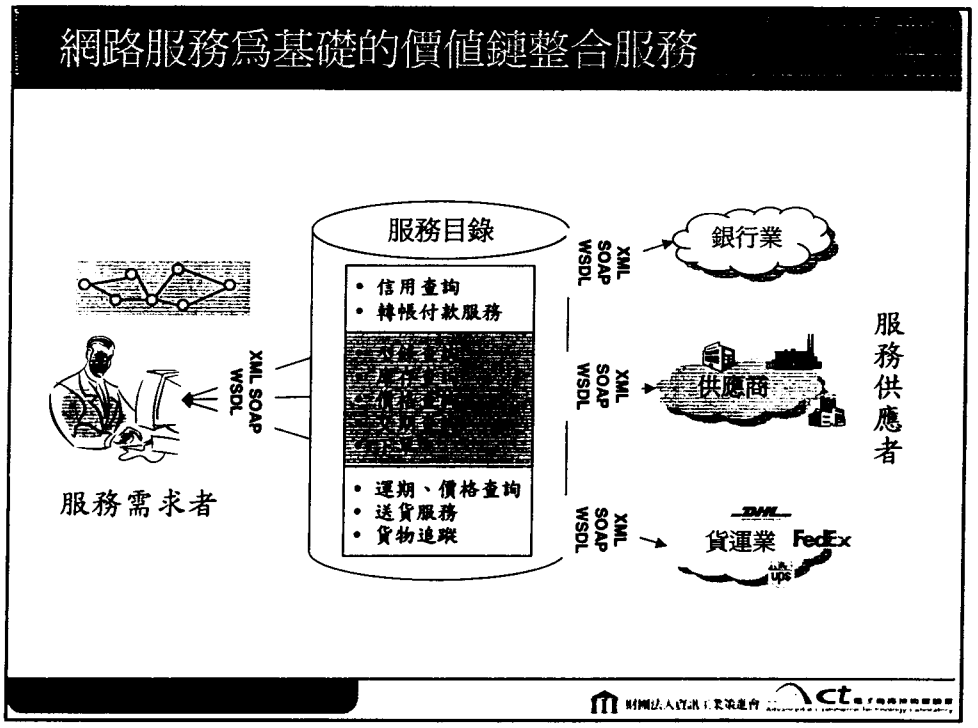


- Accessed by Multiple Devices
- Anywhere, Anytime
- Automated Discovery
- Rapid Reconfigure
- Connected by Networked Business Processes

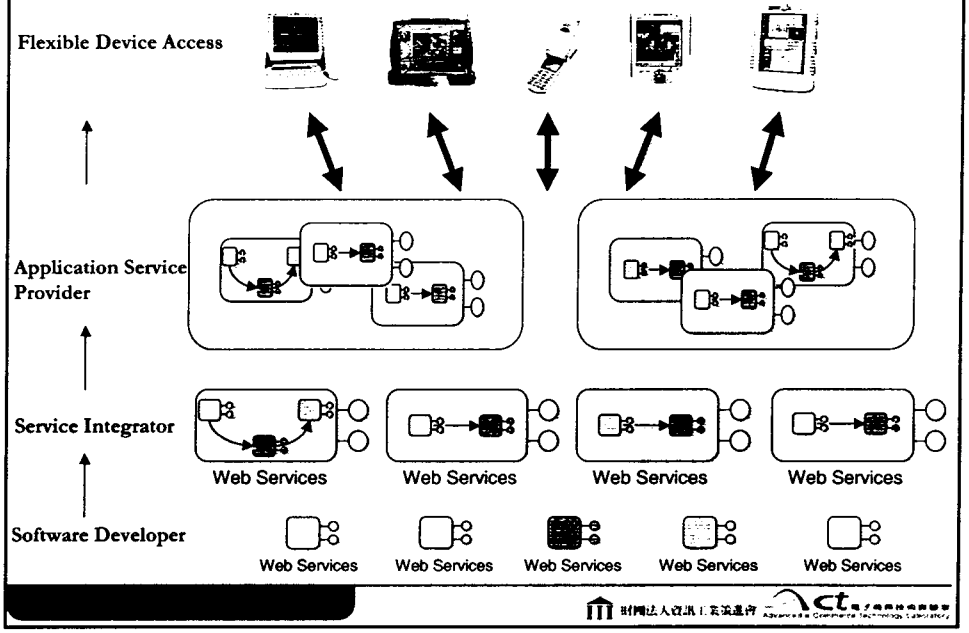

隔離、片斷的資訊系統與網路服務
CRM
Name & Address
Replication
Order Processing
Name & Address
Messaging
Logistics Company
Name & Address
Credit Card Company
Name & Address
Identity
Logon
Address
Location
Profile
Contacts
Preferences
Credit Card
Other data
Data Locked Into Applications and Devices
Similar, but Different, Data Duplicated Across Retailers, Etc
電子商場
銀行


Enterprise Application Integration (EAI)
E-Commerce System
Credit
Customer
Logistics
Billing
Logistics
Logistics
Manufacturing
Source: CBDi

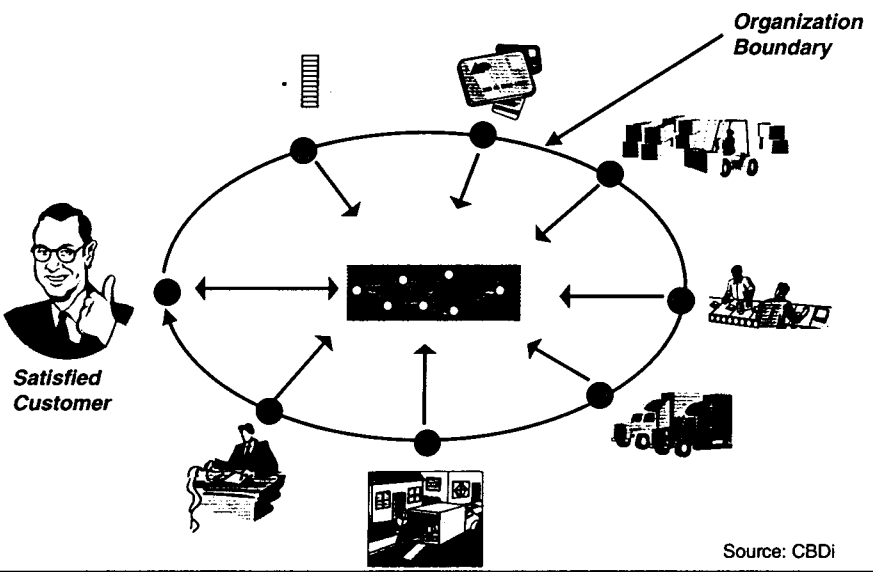

What About The Virtual Supply Chain?
E-Commerce System
Credit
• Visibility?
• Integrity?
• Efficiency?
Potential Customer
email
Logistics
Satisfied Customer
email
Manufacturing
Billing
Logistics
Logistics
Source: CBDi


網路服務爲基礎的價值鏈整合服務
服務目錄
• 信用查詢
• 轉帳付款服務
• 運期、價格查詢
• 送貨服務
• 貨物追踪
XML SOAP WSDL
銀行業
供應商
貨運業 FedEx
服務供應者
服務需求者

Web Services Integration Scenario



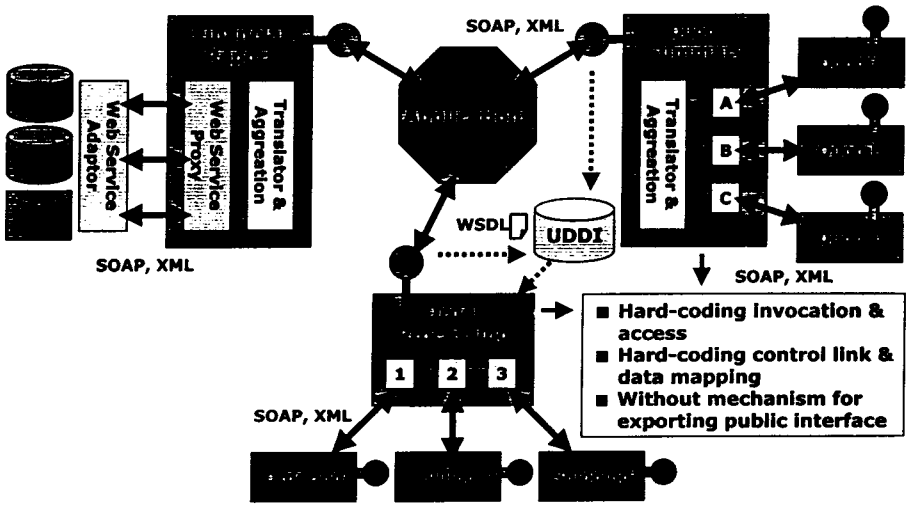
Turning Business Inside Out!



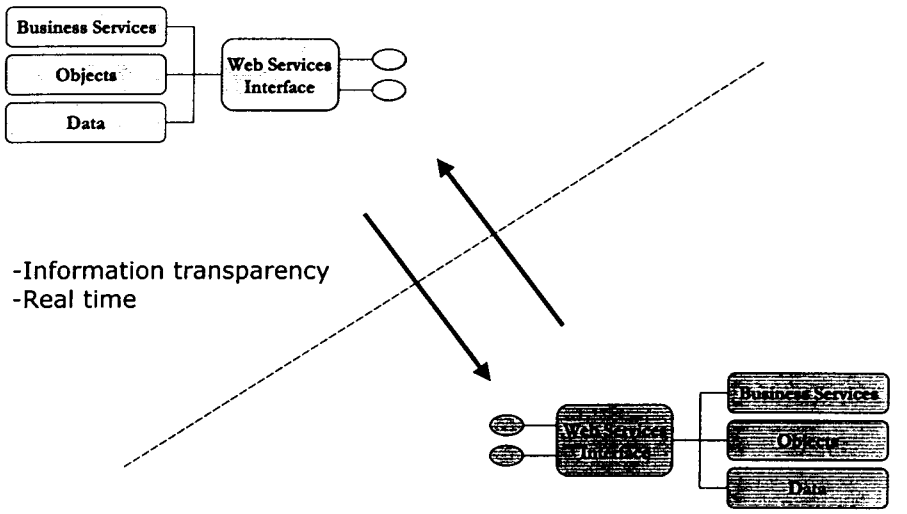
Business prospect

- Web Services Overview
  - What is Web Services?
  - Why Web Services?
  - How will Web Services reshape the business?
- **Web Services and IT application**
  - EAI
  - B2Bi
  - Content integration
  - Service integration
- Web Services Business Model
- Case Study
  - Microsoft .Net
  - IBM
  - Others

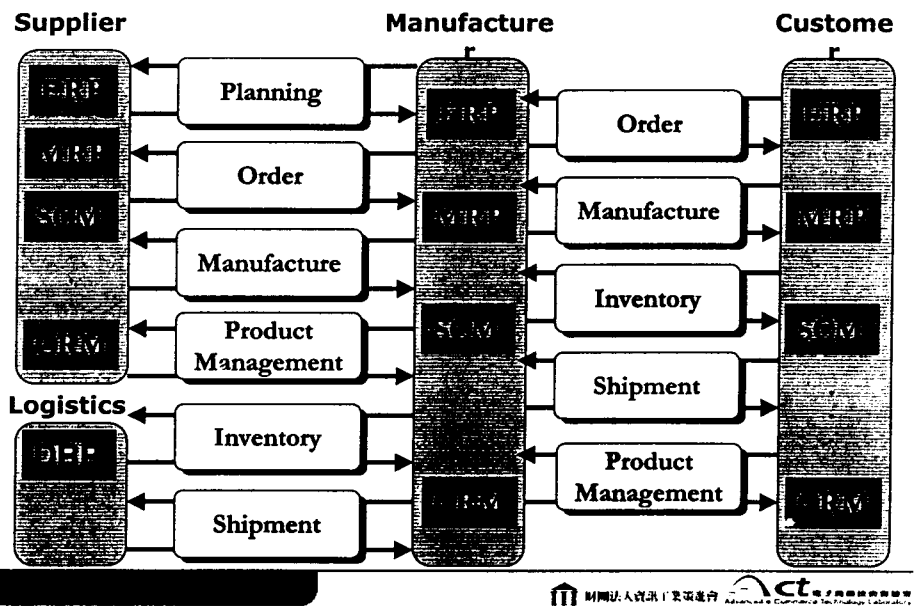
以網路服務為基礎的資料／函式整合(EAI)



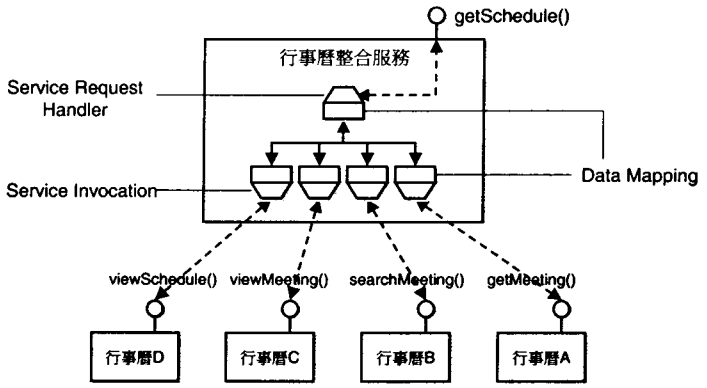
B2Bi (Data Visibility/Velocity)



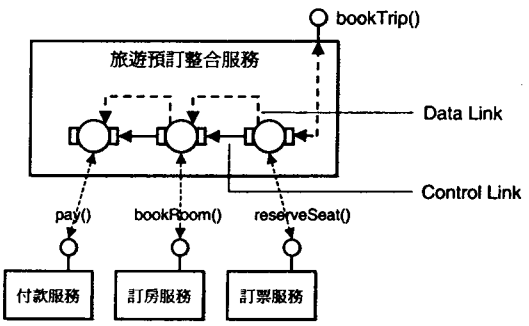
Supply Chain Visibility



Content integration



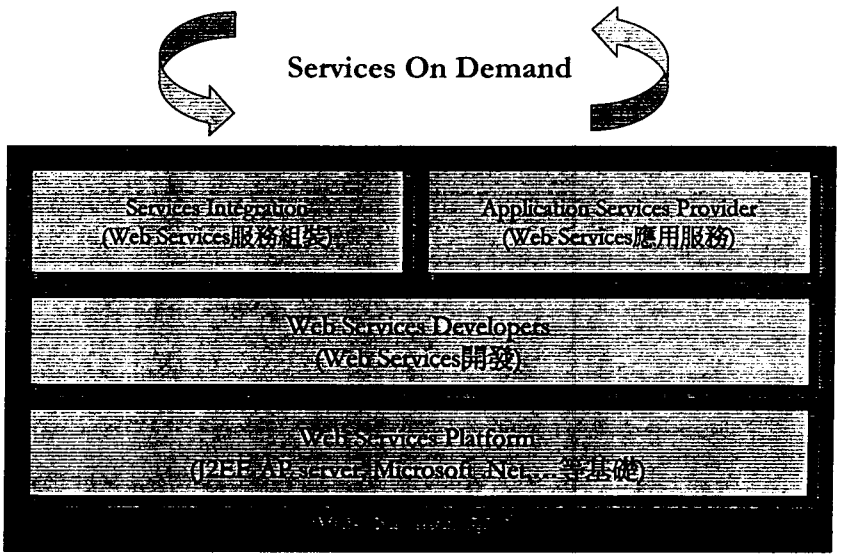
Service Integration



Business prospect

- Web Services Definition
  - What is Web Services?
  - Why Web Services?
- Web Services and IT application
  - EAI
  - B2Bi
  - Content integration
  - Service integration
- **Web Services Business Model**
- Case Study
  - Microsoft .Net
  - IBM
  - Others

Web Services對資訊服務產業的影響



Core XML Web Service Scenarios

- Data Provider (single source)
  - Stock Quote
- Data Provider (multi-source/destination)
  - Financial Service Portal
- Subscription Based Notification
  - Wireless PDA
- Maturation of Existing Applications
  - Web Service Facade
- Enterprise Application Integration
  - SAP<->Siebel<->Procurement
- Business to Business Integration
  - Linking Business Processes across domains

Extreme

Source: Microsoft

Roles in Web Service

■Business Provider
- Easy and flexible service modeling
- Security and Access control
- Transactions and sessions
- Internationalization
- Added-Value service

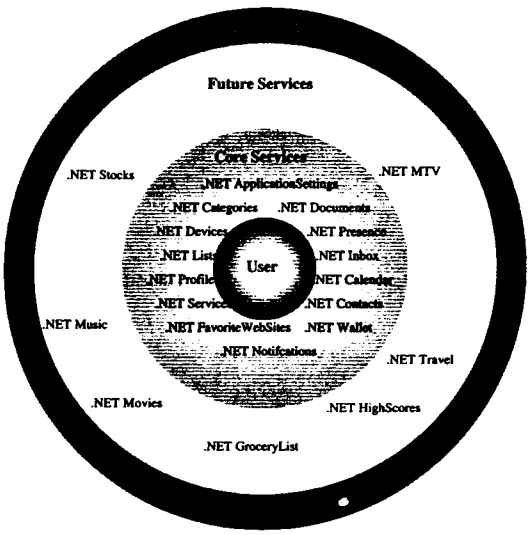
■Business Broker
- Management of Business relationships
- Invoker's Request mapping to Provider's Request
- Reliable Services
- Aggregation of services
- Flexible transport protocol for invokers
- Access control / security
- Session Management
- Business process handling / workflow
- Internationalization

■Business Invoker
- Discovery of web services
- Isolation from the Resource Providers
- Invoker defined service interfaces
- Flexible transport protocol for service invocation
- Access control / security
- Session Management
- Personalization and Internationalization

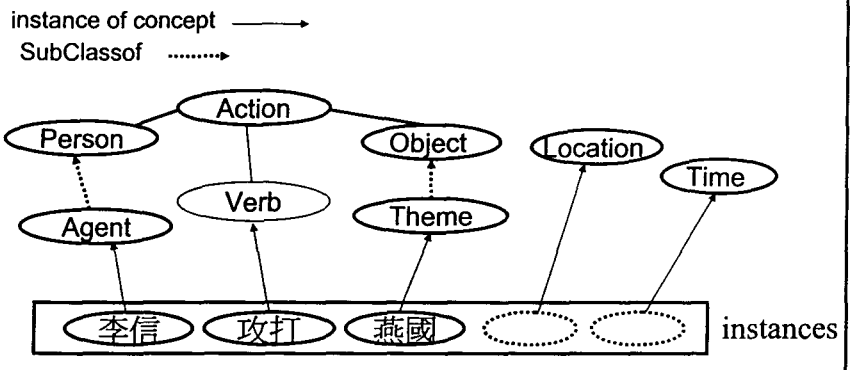
Business prospect

- Web Services Definition
  - What is Web Services?
  - Why Web Services?
- Web Services and IT application
  - EAI
  - B2Bi
  - C-Commerce
  - Business Process
- Web Services Business Model
- **Case Study**
  - Microsoft .Net
  - IBM
  - Others

Microsoft .Net My Services



Query Over Ontology



Query Over Ontology

For example

"誰攻打燕國？"

Instances are "李信 攻 燕國"

④ Even "攻打" and "攻" are not syntactically the same but is semantic meaning

④ We use query schema to recognize the meaning of users' query.

Case Study –By Microsoft

- OAG Worldwide –real time information: flight schedules and flight detail
- Clarus-A financial institution could use this web service to offer merchants or small businesses connectivity to a clearing system.
- ITopia IP-based telephony solutions turn to "Call-me-now" function enables a company to place a button on their website, via which the customer can enter their telephone number and wait to be contacted.
- Streetmap-A customer with a chain of stores can put a facility on their website that allows their visitors to enter their post code to find out where the nearest store is.
- Infobel -provides businesses with address verification & scoring services in order to reduce the risk of wrong address for EC site.
- Pantechnik -ePIX provides a single electronic interface to those carriers and offers functionality that helps shippers to manage visibility of orders as they move outbound in their distribution chain to their customers.
- FotoWire -photofinishers can use to enable customers with digital cameras to upload images from their personal computers directly to the lab to print.

Case Study –By IBM

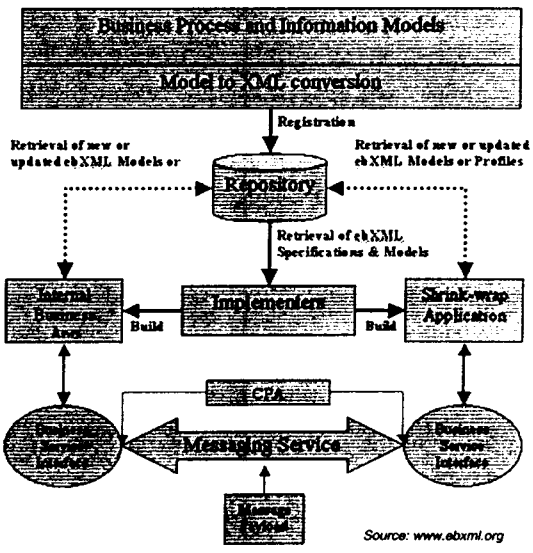
- Department of Trade and Industry, Oil & Gas Directorate - A nation's initiative: e-government
- Extend Technologies Limited - Using the Web to bring people together
- German Association of Towns and Municipalities - Federation of German towns and cities
- Hewitt Associates LLC. - Taking a global view of your employee's needs
- Hitachi Software Engineering Co., Ltd. - Integrating a disparate, proprietary world
- MicroDoc GmbH - Innovators of Web services solutions
- ORIX - A global financial services company
- Storebrand ASA - Synchronize your data

Case Study -Google

- GoogleAPIDemo is a demonstration Java app that quickly gets you searching Google, grabbing from its cache, and spell-checking
  - % java -cp googleapi.jar com.google.soap.search.GoogleAPIDemo XXXXXXXXXXXXXXXXX search rael
  - Java:
    - s.setKey(clientKey);
    - s.setQueryString(query);
    - s.setMaxResults(maxResults);
    - GoogleSearchResult r = s.doSearch();
    - GoogleSearchResultElement[] re = r.getResultElements();
    - for (int i = 0; i < re.length; i++) { System.out.println("" + re[i].getTitle() + "
"); }

ebXML – Complex Business Web Services

- Administered by United Nations and OASIS
- Enables complex business relationships to be defined
- Has adopted SOAP for transport protocol



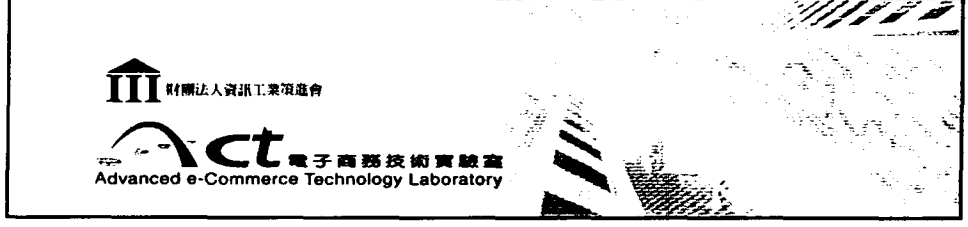
Source: www.ebxml.org

More Case Study

- www.xmethods.net
- http://www.salcentral.com/

Technical prospect

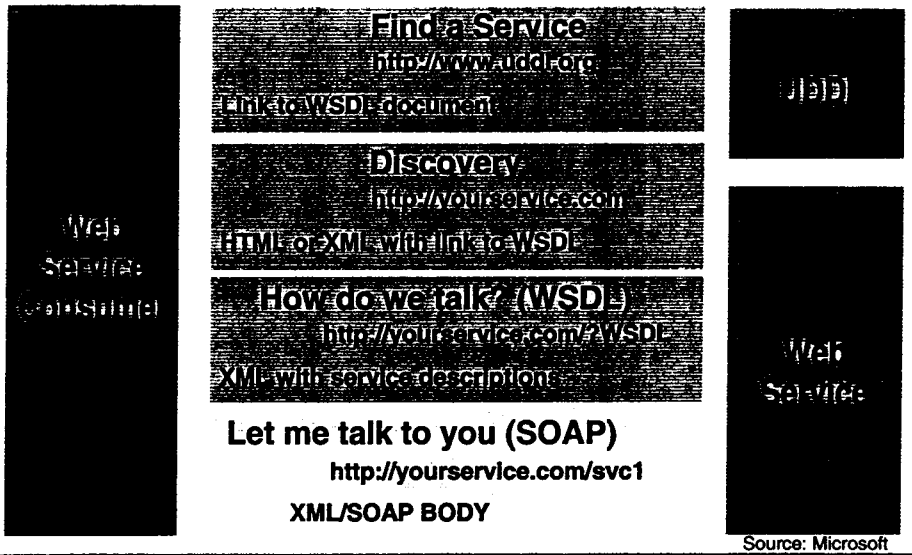
Part II



Technical

- Essential
  - SOAP (Simple Object Access Protocol)
  - WSDL (Web Service Description Language)
  - UDDI (Universal Description Discovery and Integration)
- Extension
  - Workflow/process
  - Security
  - User Interface
  - Interoperability
  - Transaction
  - QoS
  - Semantic Markup
- IT Solutions
  - Microsoft
  - IBM
  - Sun

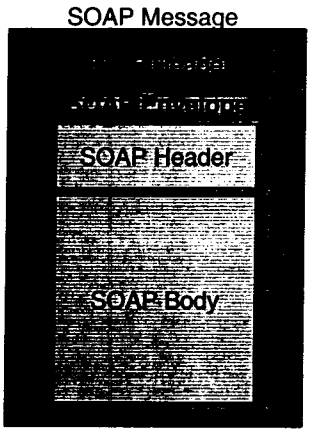
Web Services (In Practice)



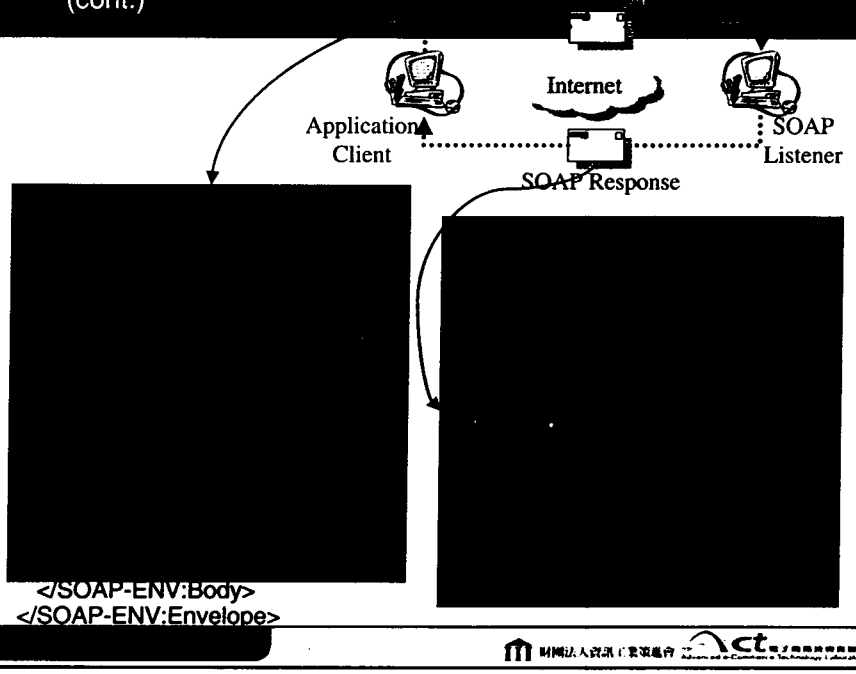
Source: Microsoft

SOAP (Simple Object Access Protocol)

- **HTTP + XML** ,simply
  - The most popular protocols on the internet
  - Firewall consideration
  - Cross platform messaging standard
  - SMTP, FTP..
  - W3C "XML Protocol"working group, Sep 13,2000
  - SOAP 1.2 is the new specification name, working draft available, Dec, 2001



SOAP (cont.)



Why SOAP will dominate

- Loosely coupled the endpoints instead of tightly coupled
  - CORBA – requires compatible ORB
  - Java RMI –requires java both
  - COM+ –requires Windows both
- Leverage existing codes instead of re-write
- W3C standards instead of proprietary
  - Promised by everyone, strongly although

WSDL (Web Service Description Language)

- WSDL is an XML format for **describing network services**
  - Binding
  - Interface
- Provide a unifying way of describing new and existing services
- The language must be open with respect to
  - service platforms (multiple type systems)
  - service interaction types: message/procedural
  - protocols: WSDL is extensible to allow new protocol bindings
- Provide the base for a Web services component composition framework

Elements in WSDL

- Implementation definition
  - Service
    - a collection of related endpoints
  - Port
    - a single endpoint defined as a combination of a binding and a network address
- Interface definition
  - Binding
    - a concrete protocol and data format specification for a particular port type
  - Operation
    - an abstract description of an action supported by the service
  - Port Type
    - an abstract set of operations supported by one or more endpoints
  - Message
    - an abstract, typed definition of the data being communicated
  - Types
    - a container for data type definitions using some type system

WSDL Example for a calculator service

```
<definitions name ='Calc' ...>
 <types>
  <schema
   targetNamespace='http://tempuri.org/type'
   />
 </types>
 <message name='Calc.Add'>
  <part name='A' type='xsd:double'/>
  <part name='B' type='xsd:double'/>
 </message>
 <message name='Calc.AddResponse'>
  <part name='Result' type='xsd:double'/>
 </message>
 <portType name='CalcSoapPort'>
  <operation name='Add'
   parameterOrder='A B'>
   <input message='wsdlns:Calc.Add' />
   <output
   message='wsdlns:Calc.AddResponse' />
  </operation>
 </portType>
```

```
 <binding name='CalcSoapBinding'
   type='wsdlns:CalcSoapPort' >
  <stk:binding .../>
  <soap:binding style='rpc'
   transport='http://schemas.xmlsoap.org/
   soap/http' />
  <operation name='Add' >
   <soap:operation
   soapAction='http://Calc.Add' />
   <input><soap:body use='encoded'
   /></input>
   <output><soap:body use='encoded'
   /></output>
  </operation>
 </binding>
 <service name='Calc' >
  <port name='CalcSoapPort'
   binding='wsdlns:CalcSoapBinding'
   >
   <soap:address
   location='http://Calc.asp' />
  </port>
 </service>
</definitions>
```

How UDDI And WSDL Work

1. 

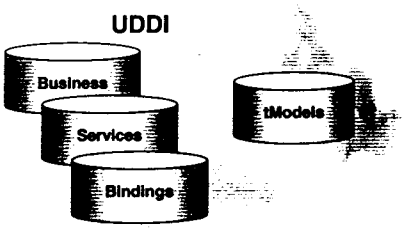
Software companies, standards bodies, and developers populate the registry with WSDL tModels.

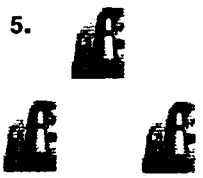


4. Marketplaces, search engines, and business applications query the registry to discover services at other companies and to facilitate integration

2. 

Implementers populate the registry with descriptions of their businesses and the services they expose



5. 

Large businesses apply the same architecture and technologies internally

Source: UDDI.org

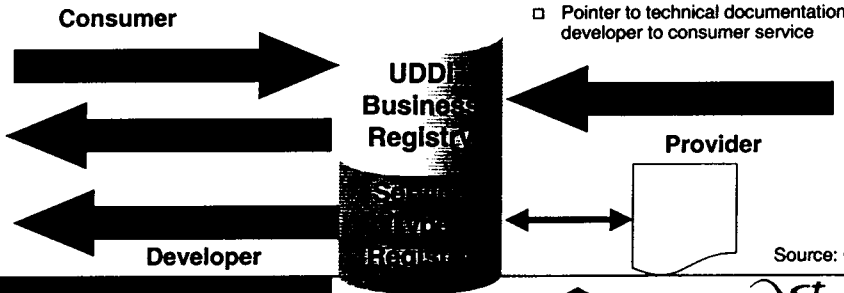
3. UDDI Registry assigns a programmatically unique [illegible] to each service and registration

Source: UDDI.org

UDDI – Universal Description, Discovery and Integration

- Provides a publication mechanism for Web Service providers
- Public Global Registry hosted by IBM, Microsoft and HP
- Administered by uddi.org
  - 260 member organizations

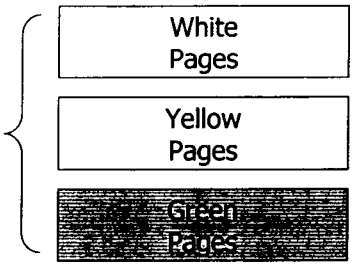
- 'White Pages'
  - General information about provider
- 'Yellow Pages'
  - Categorization of products and services
- 'Green Pages'
  - Describe service, business process, commercial, etc
- Service Type
  - Pointer to technical documentation for developer to consumer service



Source: CBDi

Registry Data

- Businesses register public information about themselves
- Standards bodies, Programmers, Businesses register information about their Service Types



Service Type Registrations

(tModel)

Source: UDDI.org

White Pages

Business name
General business description
    Any number of languages
Contact info
    Names, phone numbers,
    fax numbers, web sites, etc.
Known identifiers
    List of unique identifiers for a business
        D-U-N-S, Thomas, domain name, stock ticker symbol, other

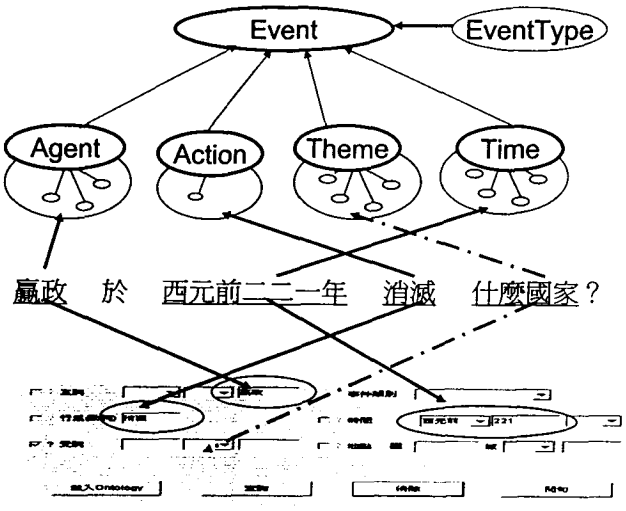
Yellow Pages

Business categories
- 3 standard taxonomies in V1
  - Industry: NAICS (Industry codes - US Govt.)
  - Product/Services: UN/SPSC (ECMA)
  - Location: Geographical taxonomy (ISO 3166)
  - Customizable after V2 version
- Implemented as name-value pairs to allow any valid taxonomy identifier to be attached to the business white page

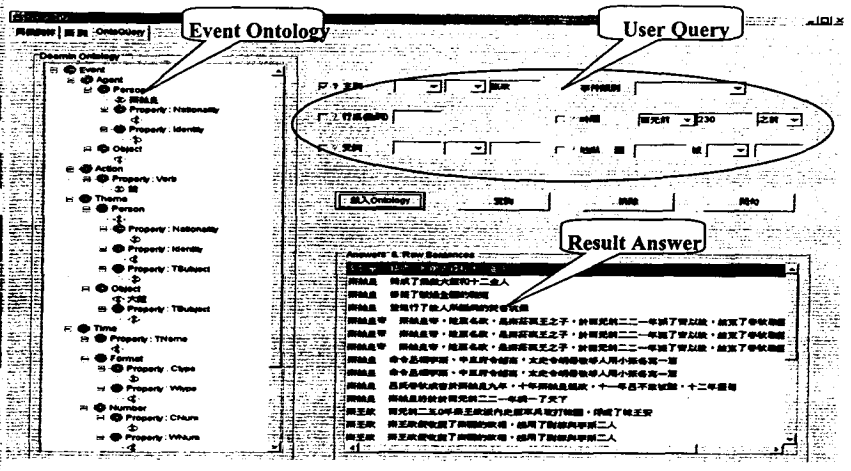
Green Pages

New set of information businesses use to describe how to "do e-commerce" with them
- Nested model
  - Business process (functional)
  - Service specifications (technical)
  - Binding information (implementation)
- Programming/platform/ implementation agnostic
- Services can also be categorized

Examples

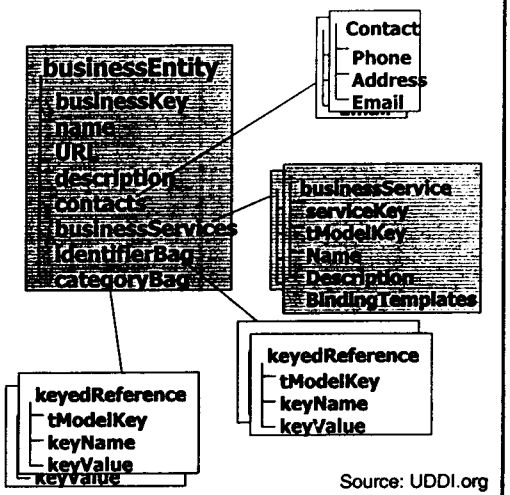


Query Interface



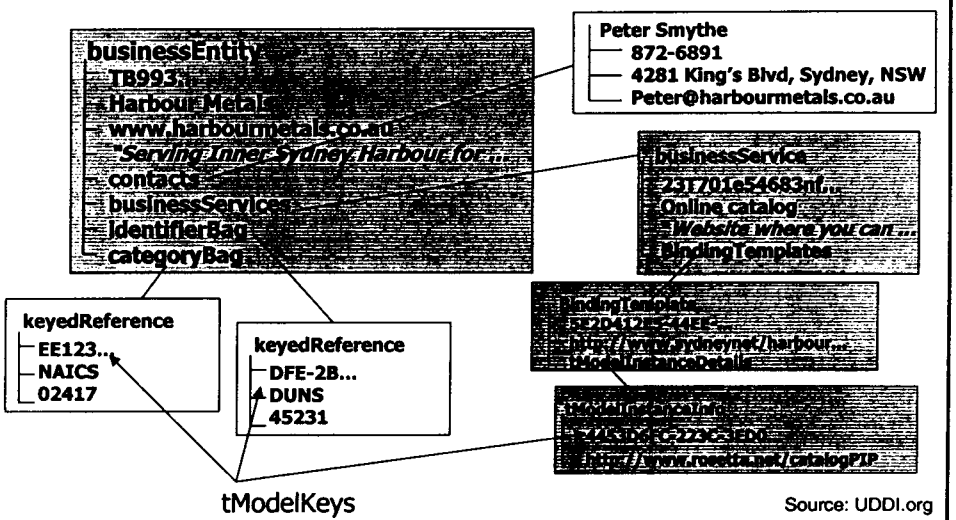
Business Registration

- XML document
- Created by end-user company (or on their behalf)
- Can have multiple service listings
- Can have multiple taxonomy listings



Source: UDDI.org

Example of a Registration



Source: UDDI.org

Registry APIs (SOAP Messages)

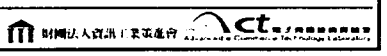
- Inquiry APIs
  - find_binding
  - find_business
  - find_relatedBusinesses
  - find_service
  - find_tModel
  - get_bindingDetail
  - get_businessDetail
  - get_businessDetailExt
  - get_serviceDetail
  - get_tModelDetail
- Publishing APIs
  - add_publisherAssertions
  - delete_binding
  - delete_business
  - delete_publisherAssertions
  - delete_service
  - delete_tModel
  - discard_authToken
  - get_assertionStatusReport
  - get_authToken
  - get_publisherAssertions
  - get_registeredInfo
  - save_binding
  - save_business
  - save_service
  - save_tModel
  - set_publisherAssertions

New in V2

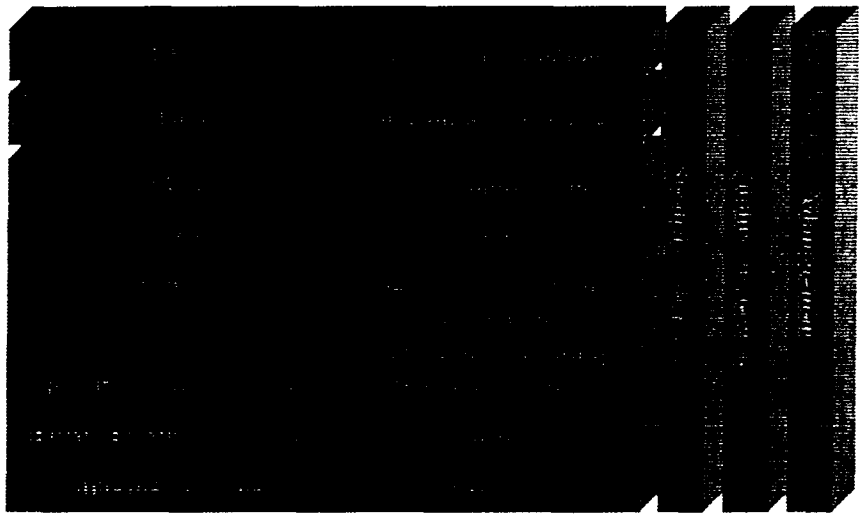


Technical

- Essential
  - SOAP (Simple Object Access Protocol)
  - WSDL (Web Service Description Language)
  - UDDI (Universal Description Discovery and Integration)
- **Extension**
  - Workflow/process
  - Security
  - User Interface
  - Interoperability
  - Transaction
  - QoS
  - Semantic Markup
- IT Solutions
  - Microsoft
  - IBM
  - Sun



Stack of Web Services Technologies



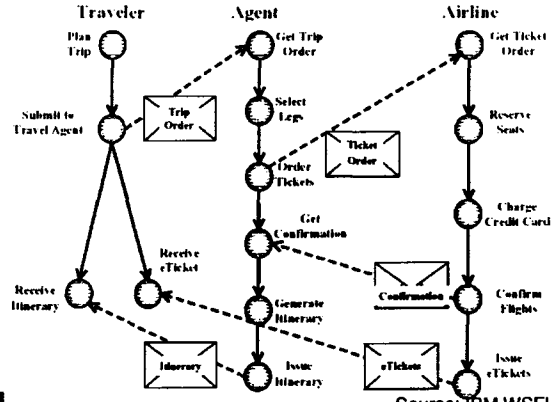
Source: IBM

Emerging standards Acronyms

Discovery	UDDI (Universal Description, Discovery and Integration) USML (UDDI Search Markup Language) WS-Inspection (Web Services Inspection Language)
Description	WSCL (Web Services Conversation Language) WSCM (Web Services Component Model) WSDL (Web Services Description Language) WSEL (Web Services Endpoint Language) WSFL (Web Services Flow Language) WSML (Web Services Meta Language) WSXL (Web Services Experience Language) WSUI (Web Services User Interface) XLANG (Web Services for Business Process Design)
Packaging and Extensions	SOAP (Simple Object Access Protocol) SOAP-DSIG (SOAP Security Extensions: Digital Signature) SWA (SOAP Messages with Attachments) WS-License (Web Services License Language) WS-Referral (Web Services Referral Protocol) WS-Routing (Web Services Routing Protocol) WS-Security (Web Services Security Language)
Transport	DIME (Direct Internet Message Encapsulation) HTTPR (Reliable HTTP)

Workflow/Process

- Workflow is the organization of processes into a well-defined flow of operations to fulfill a business need.
- To specifies the appropriate usage pattern and interaction pattern of a collection of Web Services.
- Description of Web Services compositions.
  - IBM WSFL
  - Microsoft Xlang
  - BPMI BPML



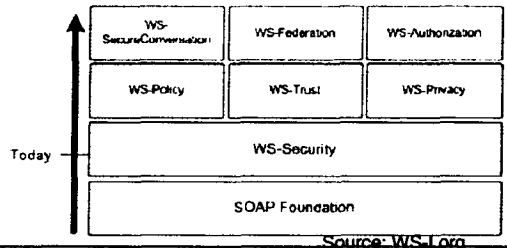
Source: IBM WSFL SPEC

Security

- Identification: I'm Alice.
- Authentication: I can prove I'm Alice.
- Authorization: Alice can read but can't write.
- Integrity: No one change this during transmission.
- Confidentiality: No one read this during transmission.
- Audit: All transactions are recorded for checking.
- non-repudiation: Yeah I can prove that you did do that to me.

WS-Security

- IBM, Microsoft and VeriSign push new standard for web services (April,2002)
- Besides WS-Security, others to come are WS-Policy, WS-Trust, WS-Privacy, WS-Secure Conversation, WS-Federation and WS-Authorization. The names are fairly self-describing. However they all are layered on top of WS-Security.
- The WS-Security specification describes how to attach signature and encryption headers to SOAP messages. In addition, it describes how to attach security tokens, including binary security tokens such as X.509 certificates and Kerberos tickets, to messages.



Source: WS-I.org

User Interface

- For coupling web services with presentation and interaction information, allowing them to be directly embedded into e-business web sites.
  - Epicenter -WSUI(Web Service User Interface)
  - IBM –WSXL (Web Services Experience Language)
  - OASIS
    - WSCM(Web Services Component Model)
    - →WSIA(Web Services for Interactive Application)
    - WSRP(Web Services for Remote Portals)(IBM)

Interoperability-WS-I

- WS-I is an open, industry organization chartered to promote Web services interoperability across platforms, operating systems, and programming languages. The organization works across the industry and standards organizations to respond to customer needs by providing guidance, best practices, and resources for developing Web services solutions.
  - The Basic Web Services Profile Working Group
  - The Sample Applications Working Group
  - The Test Materials and Tools Development Working Group

Transaction

- The interdependent workflows among multiple trading partners, which drive business transactions, need to be coordinated to ensure that the outcome of the transaction is reliable.
- including transactions that span multiple enterprises and long lasting transactions.
- ACID properties
- OASIS
  - BTP (Business Transaction Protocol)

Quality of Service (QoS)

- Non-functional properties to match the needs of service requestors with those of the service provider's based on the network resources available.
- Still stay vague, might cover:
  - Availability
  - Accessibility
  - Integrity
  - Performance
  - Reliability
  - Regulatory
  - Security
- IBM:HTTPR,WSEL(Web Services Endpoint)

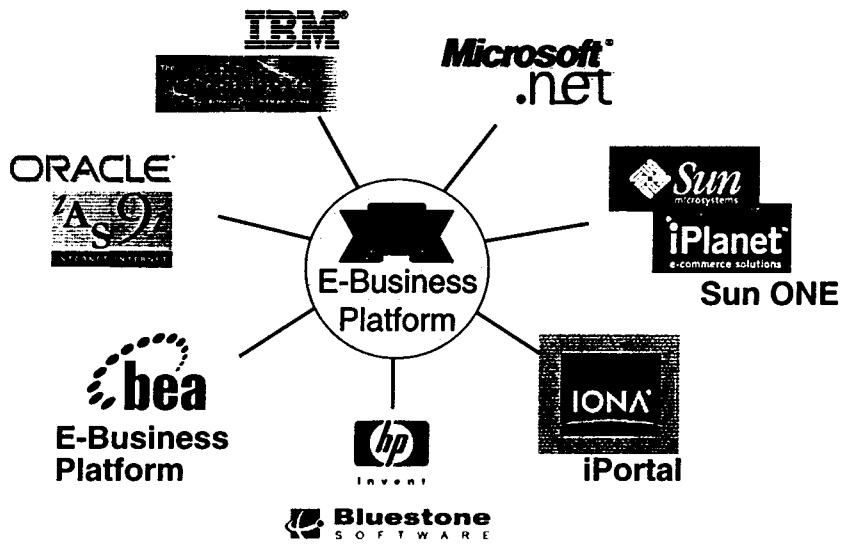
Semantic markup

- To provide a precise understanding between two party
- "Semantic Web" -Introduced by the inventor of Web, director of W3C, MIT Tim Berners-Lee.
  - A new form of Web content that is meaningful to computers
  - Expressing Meaning / Knowledge Representation
  - Ontology /Agent
- Usually refer to "call the right service ", "Service Automation"
- DAML-S(DARPA Agent Markup Language for Web Services)
- WSMF(Web Service Modeling Framework)

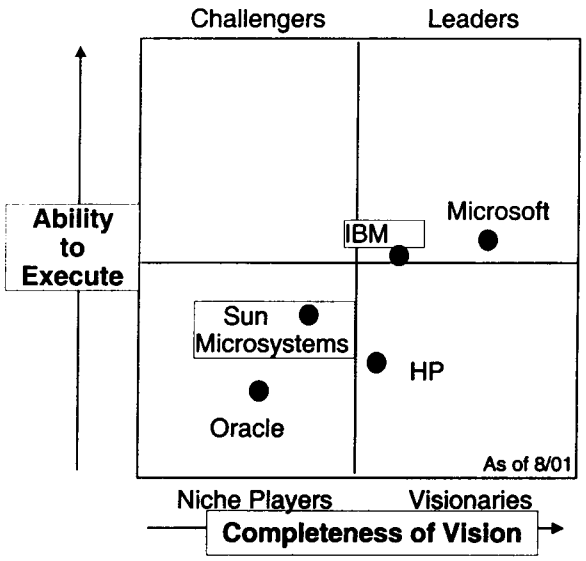
Technical

- Essential
  - SOAP (Simple Object Access Protocol)
  - WSDL (Web Service Description Language)
  - UDDI (Universal Description Discovery and Integration)
- Extension
  - Workflow/process
  - Security
  - User Interface
  - Interoperability
  - Transaction
  - QoS
  - Semantic Markup
- **IT Solutions**
  - Microsoft
  - IBM
  - Sun

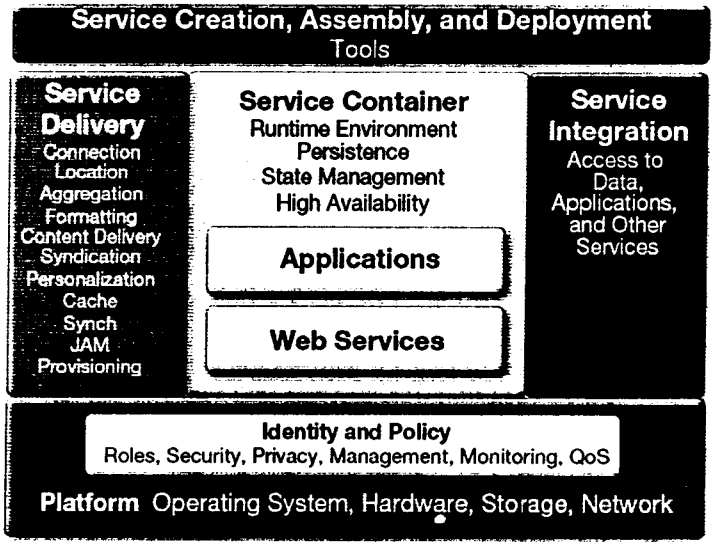
The Battle of the Platforms



Major Vendor Web Services Platform Influence

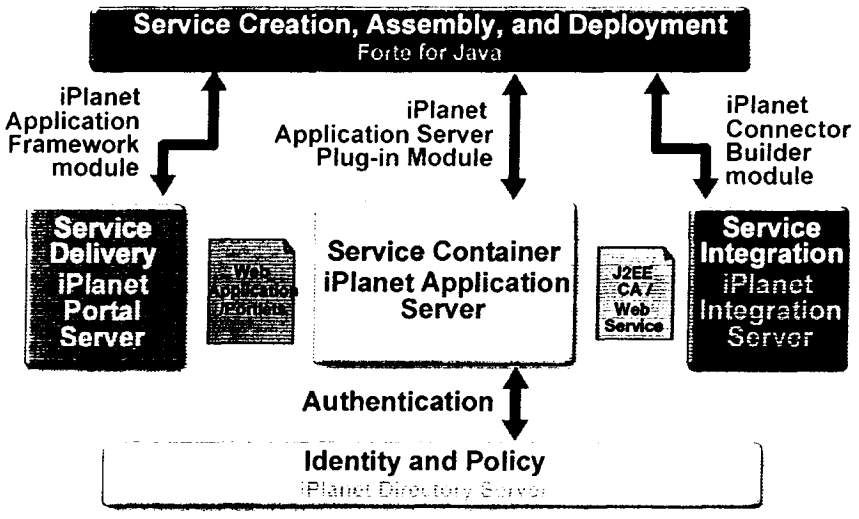


Sun ONE (Sun Open Net Environment)



Source: JavaOne 2002

Sun ONE: Application Development



Source: JavaOne 2002

SUN Forte for Java Web Services Support

- Create Web Services from EJB components
- Generate WSDL files
- Generate Java Client Proxies
- Generate Test Web Apps
- Deploy/Execute to iPlanet Application Server
- Register to UDDI from the IDE

Who-queries

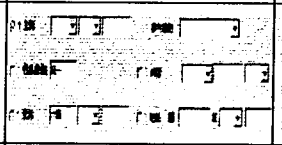
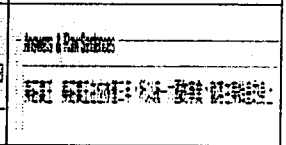
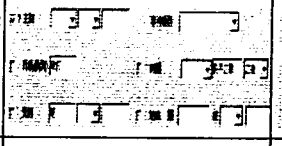
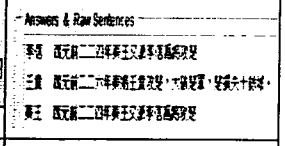
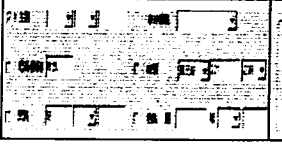
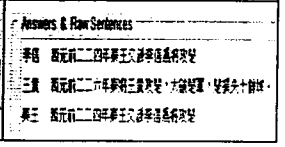
SN	Simple Question	Query Interface	Results from Interface
1	誰统一六國？		
2	長平之戰後，有哪些人攻打過楚？		
3	西元前 221 年之前，有哪些人攻伐過楚？		

Table 2 A type of questions by asking "Who"

What-queries

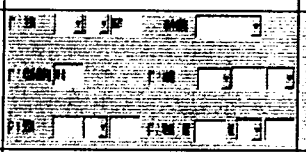
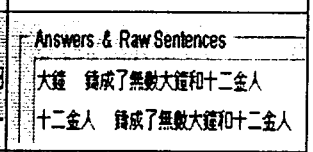
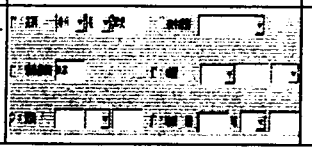
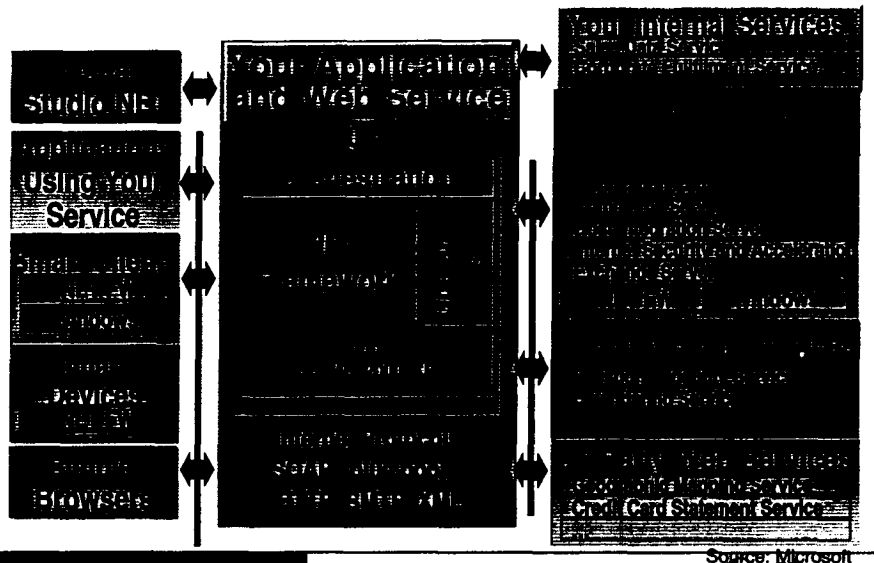
SN	Simple Question	Query Interface	Results from Interface
1	嬴政鑄造過什麼？		
2	呂不韋死後，還有戰爭事件？	N/A	N/A
3	所有官吏的執掌分別工作內容？		N/A

Table 3 A type of questions by asking "What"

Java XML Technologies

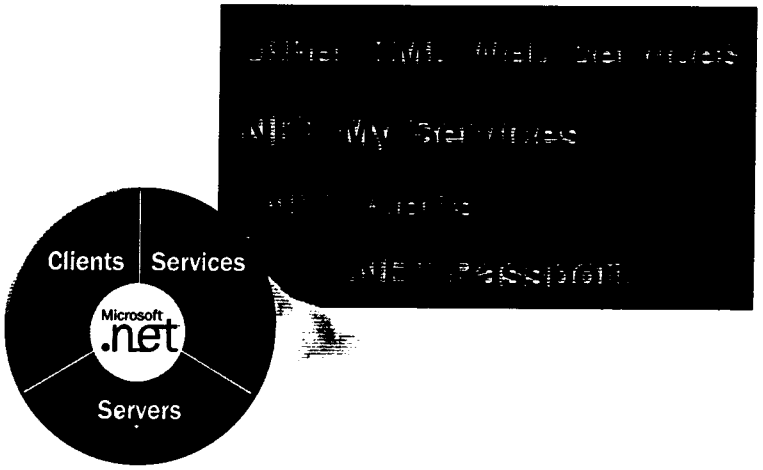
- Java XML APIs
  - Java API for XML Processing
  - Java API for XML Binding
  - Java API for XML-based RPC
  - Java API for XML Registries
  - Java API for XML Messaging
- Technology distribution
  - Java XML Pack
  - Java Web Services Developer Pack

Microsoft .NET Platform Blueprint



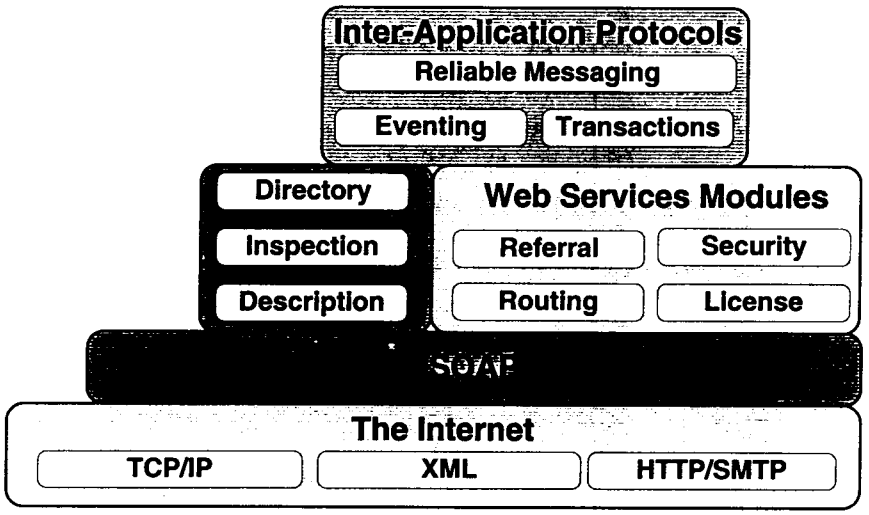
Source: Microsoft

Microsoft .Net Building Block Services



Source: Microsoft

Web Services Roadmap (Microsoft GXA)



Source: Microsoft

IBM's Web Services Toolkit V3.1

- Web Services Security specification
- BuyerSeller demonstration
- Business Explorer for Web Services (BE4WS)
- WSDL Explorer
- Web Services Management
- Utility Services
- Version 2 UDDI registry support
- AXIS (Apache SOAP 3.0)
- WSIL4J
- Windows XP support

Web Services in IBM Products

- WebSphere: SOAP and UDDI4J in next version
- DB2 Universal Database: XML Extender
  - web services for query retrieval
  - stored procedures in WSDL
- MQ Series
  - integrated into next release of the product
- WebSphere Studio Application Developer
  - wizards and tools for developing XML and Web Services applications faster and easier
- Lotus Domino:
  - enable Domino Services as Web Services
  - incorporate SOAP interfaces, XML-based messaging

Conclusions

- Web Services已成流行名詞
- 看起來已有一些可能的應用
- 但是，有待釐清的問題更多
- 也代表許許多多的發展機會
- 這是新一波軟體智財權戰爭
- 需要多點創新應用技術研究
- 並且推動成果成爲產業標準

Reference

- http://www.webservices.org
- http://www.xmethods.com
- http://www.webservicesarchitect.com
- http://java.sun.com/webservices/
- http://www-106.ibm.com/developerworks/webservices
- http://msdn.microsoft.com/webservices
- http://www.cbdiforum.com

電子化政府入口網與各機關系統整合之共通平台初探

張耀仁 博士

蕃薯藤數位科技協理

電子化政府入口網與機關網站整合共通平台初探

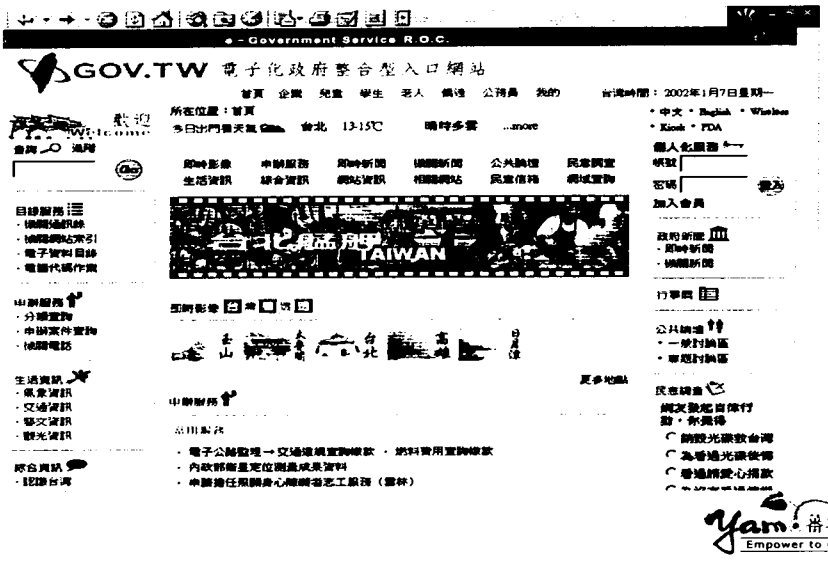
Creating Usable and Viable eGovernment Services

蕃薯藤科技 企業電子化事業處協理

張耀仁博士

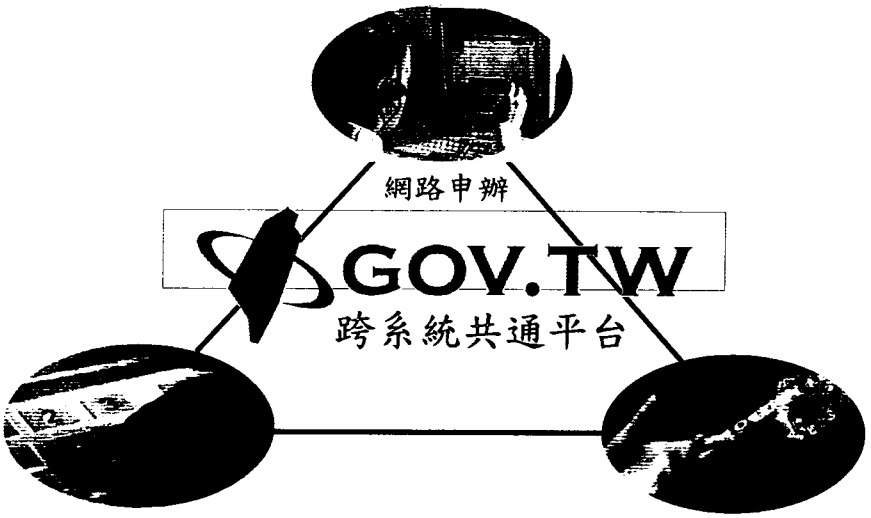
Yam 蕃薯藤 Empower to Grow

電子化政府入口網



Yam 蕃薯藤 Empower to Grow

主要服務



3

資訊查詢

- 搜尋引擎
- 目錄服務
  - 網站目錄
  - 機關通訊錄
  - 電子民意信箱目錄
- 生活資訊
  - 氣象資訊
  - 交通資訊
  - 藝文資訊
  - 觀光資訊
- 綜合資訊
  - 認識台灣
  - 政府組織
  - 重要施政
  - 專題報導
- 即時影像
  - 玉山
  - 阿里山
  - 冬山河
  - ⋮

4

雙向溝通

- 政府新聞
  - 各機關新聞
  - 即時新聞
- 機關行事曆
- 電子民意信箱
- 網路民意調查
- 公共論壇
  - 一般討論區
  - 專題討論區
- 網路社群
- 網站意見箱
- 網站最新消息

5

電子化政府入口網

- 網路上的中華民國政府
- 旗艦二子計劃，行政院研考會委外建置服務案
- 1500 項便民服務上網，5500 機關系統互連
- 實現24小時全年無休的單一窗口服務
- "政府就在你家"

Desired Goals

- Personalized end-to-end services
  - Complete services all on line
- Provide services anytime, anywhere, and on any device
- Make application services modular and easier to integrate
  - Lower cost & risk and shorter time



Criteria

- Accessibility
  - Across devices
- Scalability
  - Cluster computing and failsafe
- Stability
  - Fault isolatable
- Security
  - Based on industrial strength standards such as LDAP, PKI



IT Architectures

- Common platform
  - Maximum connectivity with common interfaces
  - Interoperability that respects the fact of existing legacies
- Open Standard
  - Of industrial strength but not vendor specific
- Federated but not centralized
  - Consistence of service description and provision across remote, distributed, heterogeneous government application systems.
  - No single point of failure (SPF)
- Self-stablizing
  - Robust, auto-recovery from isolated faults

9

Definition of Portals

- Platform for maximum interoperability and streamlined services is desired.
  - Confusion of Portal, Hub, and Gateway to be clarified
  - Experiences of Ukonline (UK), Federal.au (Australia), and FirstGov (USA) for further study and research

10

Where-queries

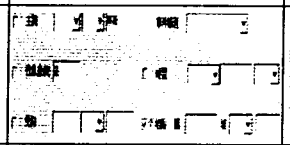
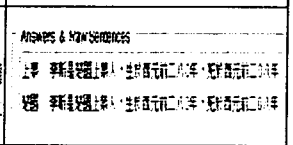
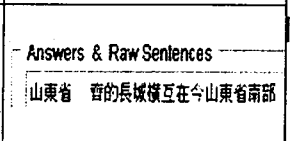
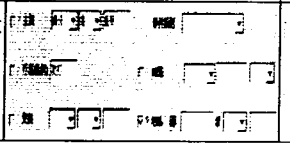
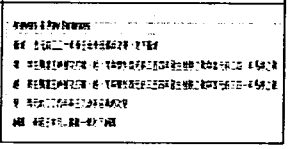
SN	Simple Question	Query Interface	Results from Interface
1	李斯是哪裡人？		
2	齊的長城在哪？		
3	戰國時期的皇帝攻打過哪些地方？		

Table 4 A type of questions by asking "Where"

When-queries

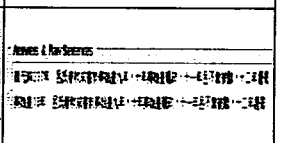
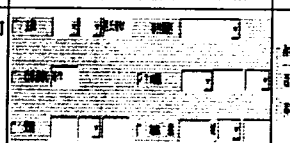
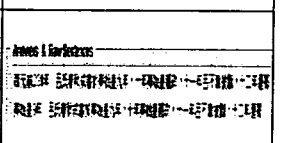
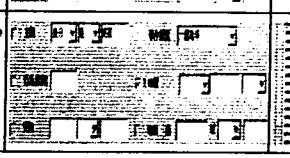
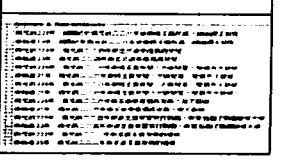
SN	Simple Question	Query Interface	Results from Interface
1	嬴政何時攝政？		
2	呂氏春秋著作於何時？		
3	戰國時期的將軍參加過戰爭的時間？		

Table 5 A type of questions by asking "When"

Portals' do and don't

- Aggregation of entries of service, not centralized factory of transaction engines
- Provision of convenience of single sign on, not mandatory gateway to accessing transactional services
- Directory information of available transactional services, not single repository or control point of business flow logics that governs transactional flow and status of those services

11

Portals' metaphors

- Factory (not good)
  - centralized factory of transaction engines
- Scoreboard (better)
  - mandatory gateway to accessing distributed transactional services
  - single repository or control point of business flow logics
- Cockpit (much better)
  - Provision of convenience of single sign on
  - Aggregation of entries of service
  - Directory information of available transactional services

12

Portals' metaphors

- "The network is the government", instead of "the portal is the government"
  - Security and transactions are distributed in nature
  - Doesn't rely on single point of security or transaction hub/gateway (therefore, no hub or gateway)

13

Beyond Portals
Integration for eGovernment

- Information Integration
  - Information consistence and integrity
- Application Integration
  - Achieve guaranteed transaction & guaranteed delivery
- Service Integration
  - Usability and viability

14

Issues to address

- Authentication & Directory architecture
- Integration of existing and new Application services
- On line payment
- Pervasive Security and Privacy
- Strategic development of information industry

15

Authentication & Directory architecture

- Authentication
  - Single sign on
  - certificate and password mechanism
- Directory architecture
  - information integration
  - mobility of preference profiles
  - repository of certificate & public key

Application service integration

- Paradigms
  - EAI (Enterprise Application Integration), Web Services
  - Heavyweight (EAI) vs. lightweight (Web Service)



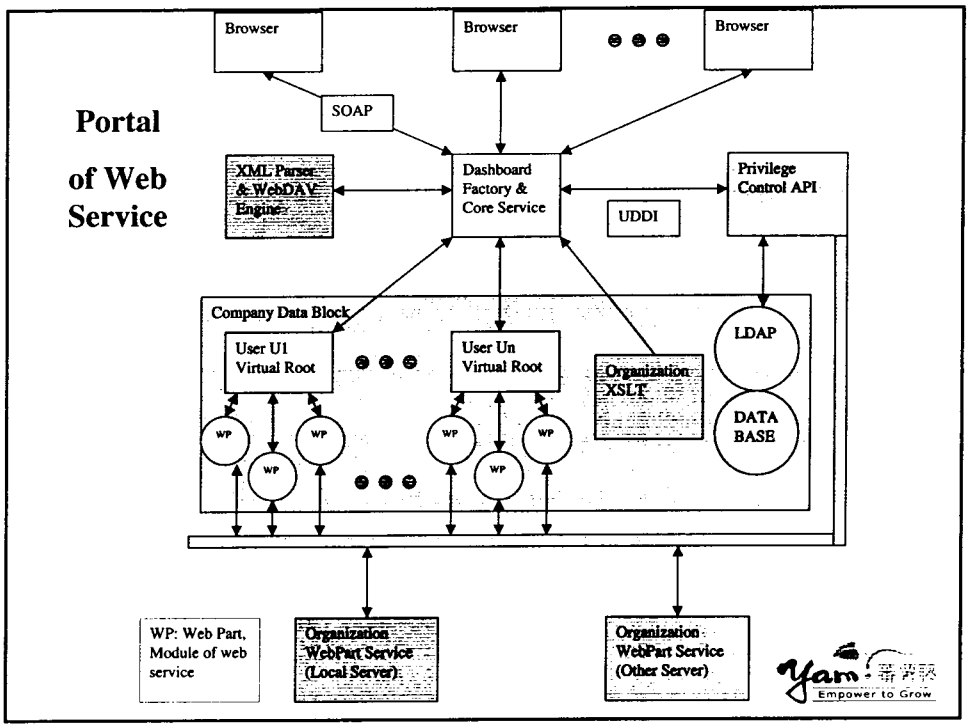
EAI

- Instruments of Remote invocation
  - Object-based: RMI, COM+, CORBA
  - Ad-Hoc: RPC, Socket Programming
- Remote messaging (Message Queue, guaranteed delivery)
- Remote 2-phase commit (Transaction Monitoring)

18

Web Services

- SOAP
  - Access to PC, Notebook, Cellular, PDA, Kiosk, IA
- UDDI
  - Consistent and universal index for users to locate on line services
- WSDL
  - Universal language and interface for definition and description of all on line services
- EAI is better for intranet-wide integration and Web Service is better for internet-wide integration of service



Pervasive Securities and Privacy

- PKI-based Digital Certificate & Payment
  - 隱密性
  - 完好性
  - 多方身分確認
  - 不可事後否認
- System and Networking Security
  - DDoS, Spoofing, Trojan Horse, etc.
  - Security holes and backdoors



On line Payment

- Payment
  - 信用卡付款
  - ATM 轉帳
  - 預付卡
  - Post-paid (ISP, Telco bill)
  - smart card (IC card)
  - micropayment
  - eCheck (used in US Navy)



交易安全機制

- 隱密性
- 完好性
- 多方身分確認
- 不可事後否認 -> 3 party handshaking.

Non- repudiation

- 3 party handshaking based on PKI
  - Non repudiation involving consumers, authorities, eGov Portal (3 parties)
  - Turnkey technology: Digital signature, Dual Encrypted Encapsulation

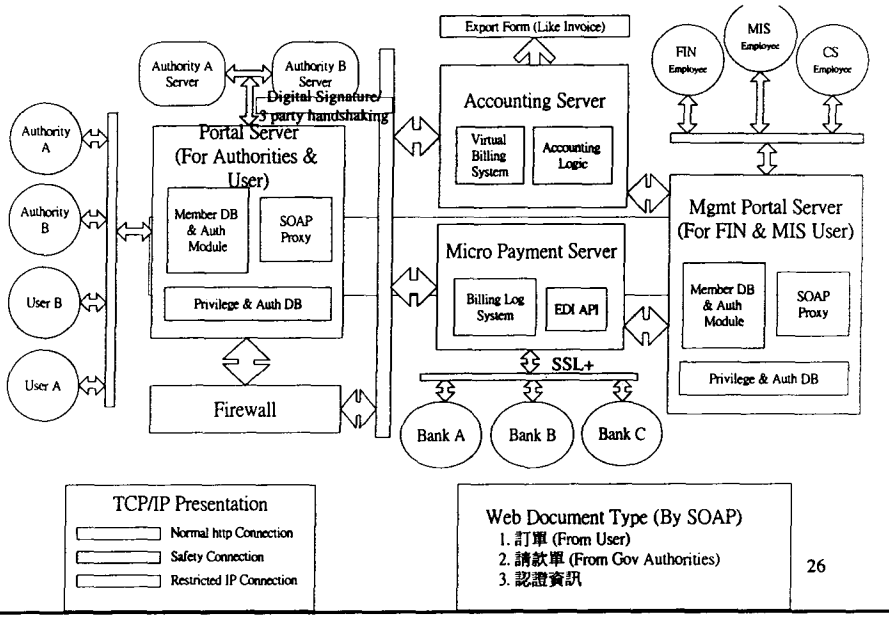
24

Non- repudiation

- Authorities cannot alter the payment info sent from users. (Payment info is encrypted by eGov public key)
- eGov Portal cannot alter the paid content item info sent from users. (Content info is encrypted by authority's public key)
- If disputes occur, User, Authorities, and eGov Portal have to present their proof of evidence and resolve the disputes. Non-repudiation is achieved by three-party handshaking.
- Confidentiality, Authentication, and Completeness are all achieved by digital certificates and SSL.

25

ePayment 共用子系統



26

Strategic Development

- In-house development vs. off-the-shelf trade-off
  - Encourage information industry to get certified internationally by participating in large-scale government funded projects
- BOT and outsourcing
  - Enable information industry to get first-hand operation experiences of large-scale application services

27

Concluding remarks

- lightweight portal
  - performs failsafe single sign on and directory lookup of government services
  - not a hub that coordinates all the transactions across all the 5500 agencies
- Web service as common, open-standard platform for interoperability
  - Light-weight, loosely-coupled, scalable, robust, and secure
- security mechanism
  - For certificate, payment, and services
  - implemented in a pervasive, structured, layered fashion
  - not bound to specific applications like before

Backup slides

29

Typical Services

Current Results

- Query types include Who, What, Where and When questions
- 55 simple historical questions
- The returned answers are 40 for correct 15 for incorrect.

Advantages

- Query Schema-Like Interface
  - split a simple question into several components by query schemas
- Using Thesaurus and Ontology
  - Deal with synonyms and different syntactical structures
- The Inference by the Relations of Concepts
  - "長平之戰後，哪些人攻打過楚？"

Scenarios-Tax filing

- Web service at personalized homepage at Feb-Mar who has not filed taxes
- Personalized tax forms
- Personalized spreadsheet
- Payment made thru assigned account
- Refund/Deposit notice thru SMS
- Refund deposited into account

- Similar Applications: License renewal

Scenarios- Passport Issuance

- Web Service for passport application
- Web Service checks police, IRS, local and national courts, DoD
- Certificate issued into IC Card
- Identify with IC Card, pick up passport at Post Office thru print-on-demand with photo shot on site

Scenarios- Pay parking fines

- Personalized web service collecting parking tickets across geographical areas
- Payment made thru default account with authorization



Scenarios- Newborns

- Hospital files birth certificate to public health agency
- Personalized web service showing all the benefits entitled to and, transferrable to bank accounts
- Birth certificate issued on line so that insurance can be purchase on line.



Web Service Essentials



Web vs. Web Service

- Why the web existing today is still not web service? Reason 1
  - Existing web provides merchant centric service to customers (Server dictates service provision, even the screen resolution)
  - Web Services are customers oriented (customized on the fly in a way users prefer)



Web vs. Web Service

- Why the web existing today is still not web service? Reason 2
  - Existing web enhances GUI for man-machine interactions
  - Web Services harmonizes man-man (woman-woman, if you like) collaboration and interaction.



Concluding Remarks- Key Functions

- Identification and Authorization
- Interoperability for G-G,G-B,G-C
- Secure Transactions on government services
- Humanity Design for Minorities
- Personalized Services for civilians and government employees
- Universal Information Retrieval



Weakness

- Erroneous Linguistic Analysis
  - "秦莊襄王在位亦僅三年，所以統一六國的事業，就落在秦始皇的身上"
  - An inverted sentence
    "掌管帝室財務的少府"
- Ontology Incompleteness
  - "呂不韋死後，還有戰爭事件？"
  - "秦的將軍有誰？"

Conclusions

- Agents require domain knowledge to retrieve and extract information
- Building sharable ontology will ensure information agents to interpret domain information in the right context and semantics
- Semantic web concepts provide a feasible environment for various agents to behave and share and exchange knowledge with each other

Conclusions

- We design a framework that can retrieve annotated information using sharable domain ontology and thesaurus.
  - The sharable domain ontology in RDF schemas.
  - A query parser that parses NL queries into query schemas in terms of XML format.
  - Tools for annotating the information into RDF instances.
  - Tools for augmenting a Chinese thesaurus of general domain with lexical items.
  - Heuristic algorithms to match the RDF queries with annotated images and documents.

ACKNOWLEDGMENT

Colleagues

- National Tsing Hua University, Taiwan
  - Von-Wun Soo,
  - Chen-Yu Lee,
  - Chao-Ming Lin
  - Chao-Chun Yeh
- National Cheng-Chih University, Taiwan
  - Jih-Shane Liu
- Simmons College, USA
  - Ching-Chih Chen

GRNATS

- MOE Programs of promoting academic excellence of universities ; project number 89-E-FA04-1-4
- NSC International Digital Library project (IDLP) NSC 90-2750-H-002-734

  (in collaboration with US NSF Chinese Memory Net project)

An Introduction to Semantic Web Portal

Ching-Long Yeh 葉慶隆
Department of Computer Science and Engineering
Tatung University
chingyeh@cse.ttu.edu.tw
http://www.cse.ttu.edu.tw/chingyeh

Outline

- A brief description to the Semantic Web Portal
- Architectural basis of the Semantic Web
- KA2: an ontology-based web portal
- An architecture of RDF triple data store
- Conclusions

Introduction to Semantic Web

- Facilities to put machine-understandable data on the Web are becoming a high priority for many communities.
- The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people.
- For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently.

Introduction to Semantic Web

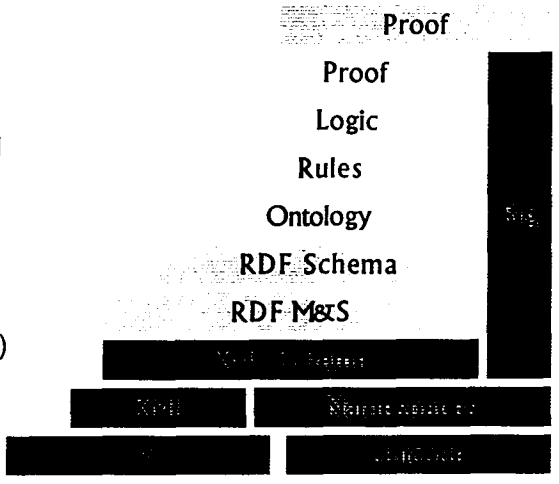
- The Semantic Web is a vision:

  the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications

- *See "W3C Semantic Web Activity," by Marja-Riitta Koivunen, for more descriptions.*

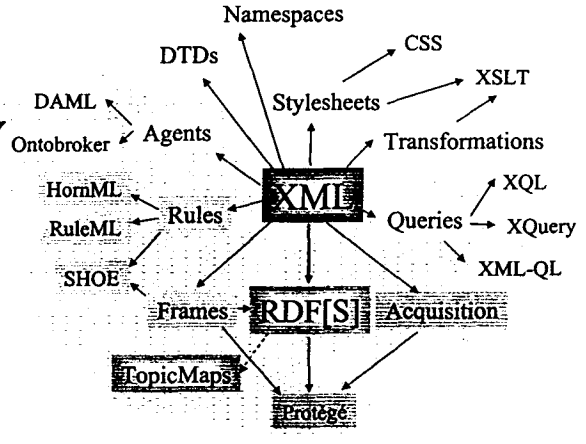
The Semantic Web Layered Architecture

Tim Berners-Lee:
"Axioms, Architecture and Aspirations"
W3C all-working group plenary Meeting
28 February 2001
(http://www.w3.org/2001/Talks/0228-tbl/slide5-0.html)

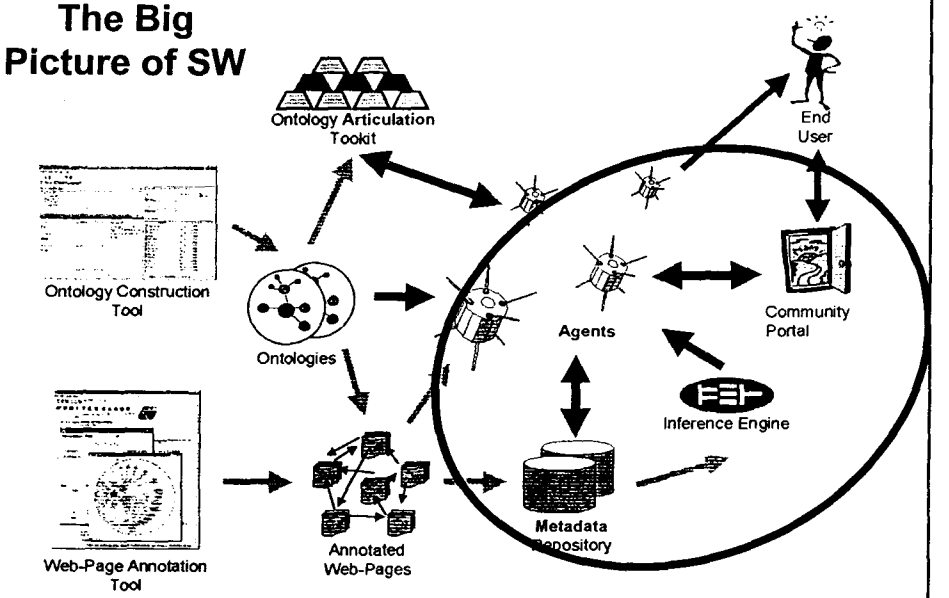


AI's Chance

- Increasing demand for formalized knowledge on the Web: AI's chance!
- XML- & RDF-based markup languages provide a 'universal' storage/interchange format for such Web-distributed knowledge representation.



The Big Picture of SW



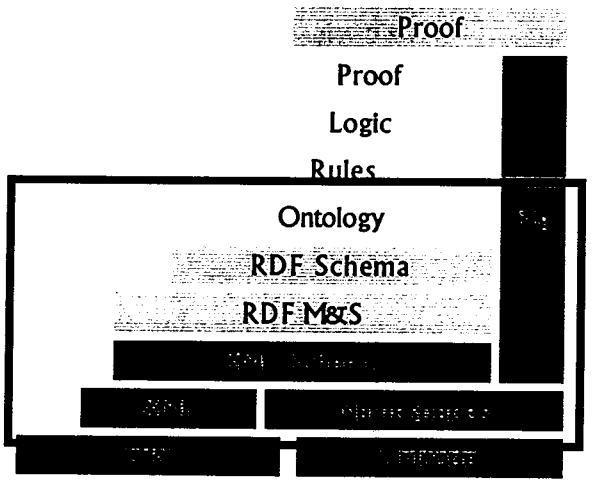
Web Portals

- A web portal is a web site that provides information content on a common topic.
  - General portals, e.g., Yahoo, Excite, Netscape, Lycos, CNET, MSN, and AOL.com
  - Specialized portal e.g., gardeners.com, semanticweb.org
- Making valuable information to be found
  - directory service,
  - search facility
  - news, e-mail,
  - community forum

Ontology-Based Web Portals

- Ontology represents
  - common knowledge and interests sharing within their community
- Tasks that ontology can be used to support a portal
  - Accessing a portal
    - Conceptual search and navigation
    - Inference capabilities
  - Providing information
    - Methods and tools accounting for the diversity of information sources

Introduction to XML-Based Ontology



What is XML?

- Extensible Markup Language
- A Syntax for Documents
- A Meta-Markup Language
- A Structural and Semantic Language, not a Formatting Language
- Not just for Web pages

Namespace Bindings

- Prefixes are bound to namespace URIs by attaching an xmlns:*prefix* attribute to the prefixed element or one of its ancestors, *prefix*:$name_1$,..., *prefix*:$name_n$
- The value of the xmlns:*prefix* attribute is a URI, which may or (unlike for DTDs!) may not point to a description of the namespace's syntax
- An element can use bindings for multiple name-spaces via attributes xmlns:$prefix_1$,..., xmlns:$prefix_m$

XML Document Instance and DTD

```
<mail:address xmlns:mail="http://www.deutschepost.de/"
        xmlns:tele="http://www.telekom.de/">
  <mail:name>Xaver M. Linde</mail:name>
  <mail:street>Wikingerufer 7</mail:street>
  <mail:town>10555 Berlin</mail:town>
  <mail:bill>12.50</mail:bill>
  <tele:phone>030/1234567</tele:phone>
  <tele:phone>030/1234568</tele:phone>
  <tele:fax>030/1234569</tele:fax>
  <tele:bill>76.20</tele:bill>
</ mail:address>
```

```
<!ELEMENT address (name, street, town,
                   bill, phone+, fax+, bill)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT street (#PCDATA)>
<!ELEMENT town (#PCDATA)>
<!ELEMENT bill (#PCDATA)>
<!ELEMENT phone (#PCDATA)>
<!ELEMENT fax (#PCDATA)>
<!ELEMENT ARTIST (#PCDATA)>
```

XML Schema

- An XML syntax that's an alternative and/or supplement to DTDs
- Data typing of element and attribute content

RDF M&S

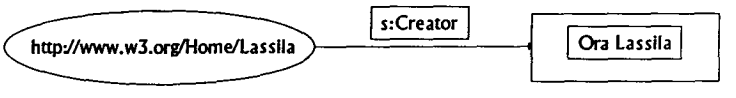
- RDF (Resource Description Framework)
  - Beyond Machine readable to *Machine understandable*
- RDF consists of two parts
  - RDF Model (a set of triples)
  - RDF Syntax (different XML serialization syntaxes)
- RDF Schema for definition of Vocabularies (simple Ontologies) for RDF (and in RDF)

RDF Data Model

- **Resources**
  - A resource is a thing you talk about (can reference)
  - Resources have URI's
  - RDF definitions are themselves Resources (linkage, see requirement 1)
- **Properties**
  - slots, define relationships to other resources or atomic values
- **Statements**
  - "Resource has Property with Value"
  - (Values can be resources or atomic XML data)
- **Similar to Frame Systems**

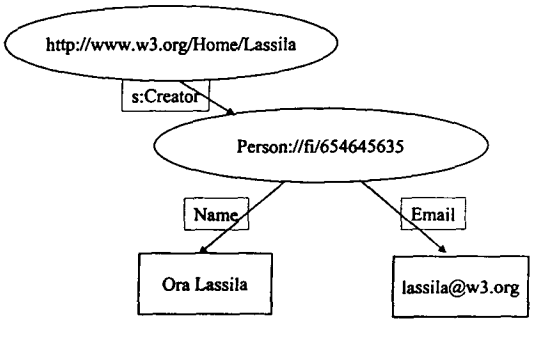
A Simple Example

- **Statement**
  - "Ora Lassila is the creator of the resource http://www.w3.org/Home/Lassila"
- **Structure**
  - Resource (subject) http://www.w3.org/Home/Lassila
  - Property (predicate) http://www.schema.org/#Creator
  - Value (object) "Ora Lassila"
- **Directed graph**



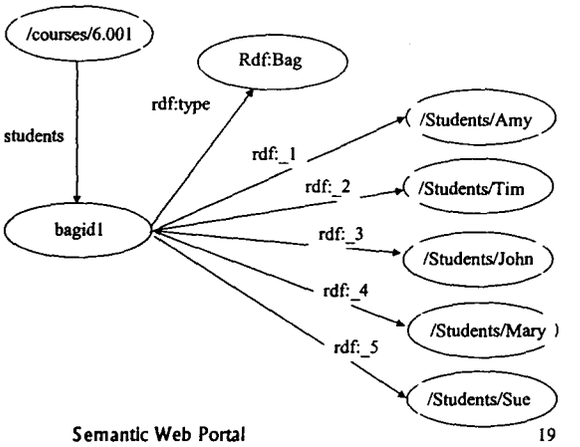
Another Example

- To add properties to Creator, point through an intermediate Resource.



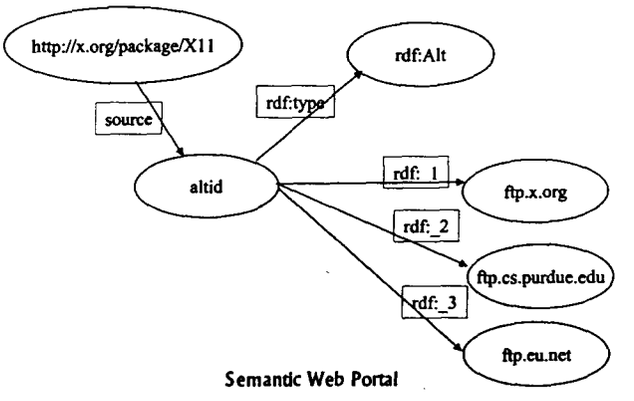
Example: Bag

- *The students in course 6.001 are Amy, Tim, John, Mary, and Sue*



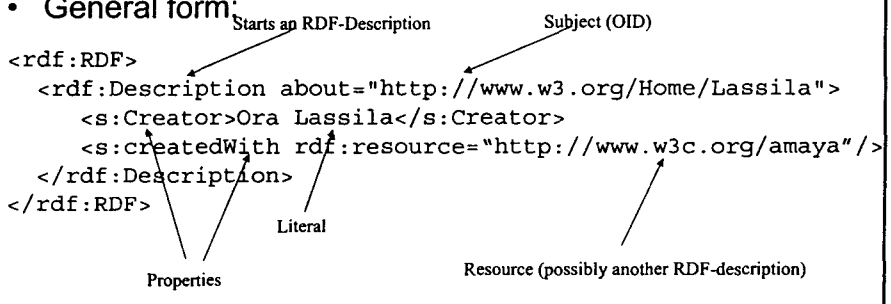
Example: Alternative

- *The source code for X11 may be found at ftp.x.org, ftp.cs.purdue.edu, or ftp.eu.net*



RDF Syntax I

- Data model does not enforce particular syntax
- Specification suggests many different syntaxes based on XML
- General form:

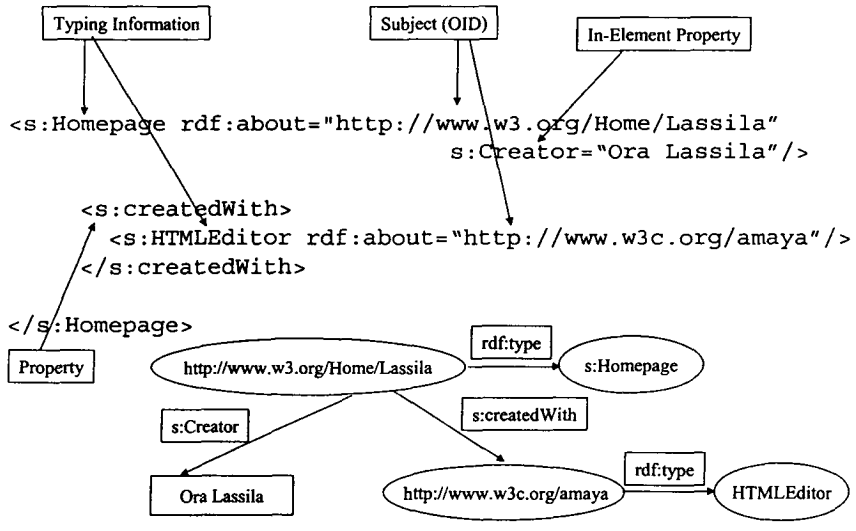


Resulting Graph




```
<rdf:RDF>
 <rdf:Description about="http://www.w3.org/Home/Lassila">
   <s:Creator>Ora Lassila</s:Creator>
   <s:createdWith rdf:resource="http://www.w3c.org/amaya"/>
 </rdf:Description>
</rdf:RDF>
```

RDF Syntax II: Syntactic Varieties



RDF Schema (RDFS)

- RDF just defines the data model
- Need for definition of vocabularies for the data model - an Ontology Language!
- The RDF Schema mechanism provides a basic *type system* for use in RDF models.
- The RDF schema specification language is less expressive, but much simpler to implement, than full predicate calculus languages such as CycL and KIF.

Most Important Modeling Primitives

- Core Classes
  - Root-Class `rdfs:Resource`
  - MetaClass `rdfs:Class`
  - Literals `rdfs:Literal`
- `rdfs:subclassOf`-property
- Inherited from RDF: properties (slots)
- `rdfs:domain` & `rdfs:range`
- `rdfs:label`, `rdfs:comment`, etc.
- Inherited from RDF: InstanceOf (`rdf:type`)

DAML+OIL: an Ontology Language

- Extension of RDF Schema
- Ontology Language DAML+OIL: Result of a Joint (European + US-American) Committee
- Extension of RDF Schema
  - Class Expressions (Intersection, Union, Complement)
  - XML Schema Datatypes
  - Enumerations
  - Property Restrictions
    - Cardinality Constraints
    - Value Restrictions

Example: Intersection & Synonyms

```
<daml:Class rdf:ID="TallMan">
  <daml:intersectionOf rdf:parseType="daml:collection">
     <daml:Class rdf:about="#TallThing"/>
     <daml:Class rdf:about="#Man"/>
  </daml:intersectionOf>
</daml:Class>

<daml:Class rdf:ID="HumanBeing">
  <daml:sameClassAs rdf:resource="#Person"/>
</daml:Class>
```


Example: Disjoint & Complement

```
<daml:Disjoint rdf:parseType="daml:collection">
  <daml:Class rdf:about="#Car"/>
  <daml:Class rdf:about="#Person"/>
  <daml:Class rdf:about="#Plant"/>
</daml:Disjoint>
```

`Disjoint` not strictly necessary, since expressible via pairwise `subClassOf` of `complementOf`, as for `Car` and `Person`:

```
<daml:Class rdf:ID="Car">
  <rdfs:comment>no car is a person</rdfs:comment>
  <rdfs:subClassOf>
     <daml:Class>
       <daml:complementOf rdf:resource="#Person"/>
     </daml:Class>
  </rdfs:subClassOf>
</daml:Class>
```

Example: Properties

(Transitive, Inverse, subProperty, UniqueProperty, range, Datatypes)

```
<daml:TransitiveProperty rdf:ID="hasAncestor"/>

<daml:ObjectProperty rdf:ID="hasChild">
  <daml:inverseOf rdf:resource="#hasParent"/>
</daml:ObjectProperty>

<daml:UniqueProperty rdf:ID="hasMother">
  <rdfs:subPropertyOf rdf:resource="#hasParent"/>
  <rdfs:range rdf:resource="#Female"/>
</daml:UniqueProperty>

<daml:DatatypeProperty rdf:ID="age">
  <rdf:type rdf:resource="http://www.daml.org/2001/03/daml+oil#UniqueProperty"/>
  <rdfs:range rdf:resource="http://www.w3.org/.../XMLSchema#nonNegativeInteger"/>
</daml:DatatypeProperty>
```


Using User-defined Datatypes (based on XML Schema)

```
<xsd:simpleType name="over17">
  <!--over17 is an XMLS datatype based on decimal-->
  <!--with the added restriction that values must be >=18-->
  <xsd:restriction base="xsd:decimal">
    <xsd:minInclusive value="18"/>
  </xsd:restriction>
</xsd:simpleType>

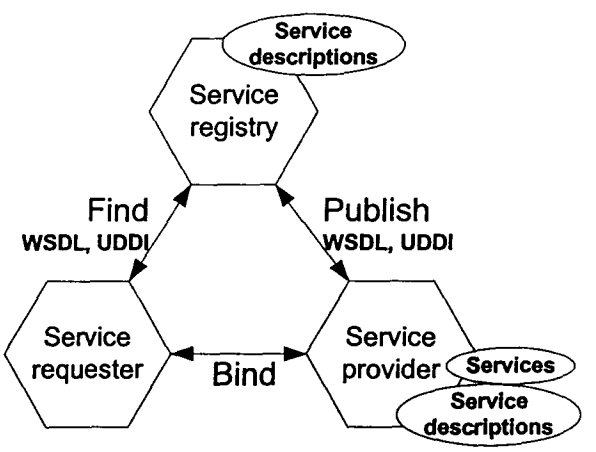
<daml:Class rdf:ID="Adult">
  <daml:intersectionOf rdf:parseType="daml:collection">
    <daml:Class rdf:about="#Person"/>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#age"/>
      <daml:hasClass rdf:resource="somefile#over17"/>
    </daml:Restriction>
  </daml:intersectionOf>
</daml:Class>
```

Instances (Individuals)

```
<daml:Class rdf:ID="Person">
  . . .
</daml:Class>


<Person rdf:ID="Adam">
  <rdfs:label>Adam</rdfs:label>
  <rdfs:comment>Adam is a person.</rdfs:comment>
  <age><xsd:integer rdf:value="13"/></age>
  <shoesize>
     <xsd:decimal rdf:value="9.5"/>
  </shoesize>
</Person>
```

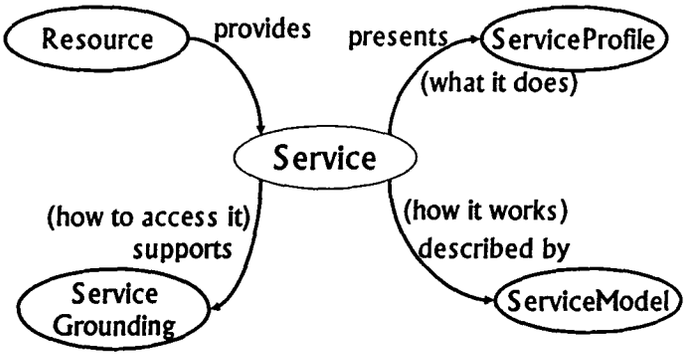

Web Services



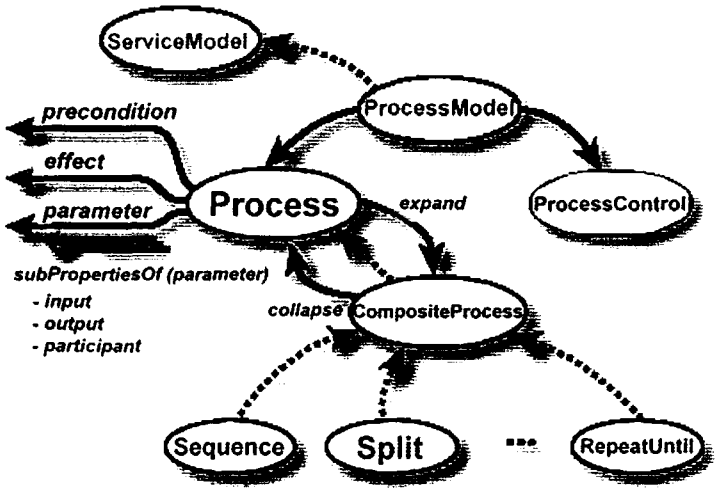
DAML-S

- Users and software agents should be able to discover, invoke, compose, and monitor Web resources offering particular services and having particular properties.
- As part of the DARPA Agent Markup Language program, we have begun to develop an ontology of services, called DAML-S.

Top Level of the Service Ontology

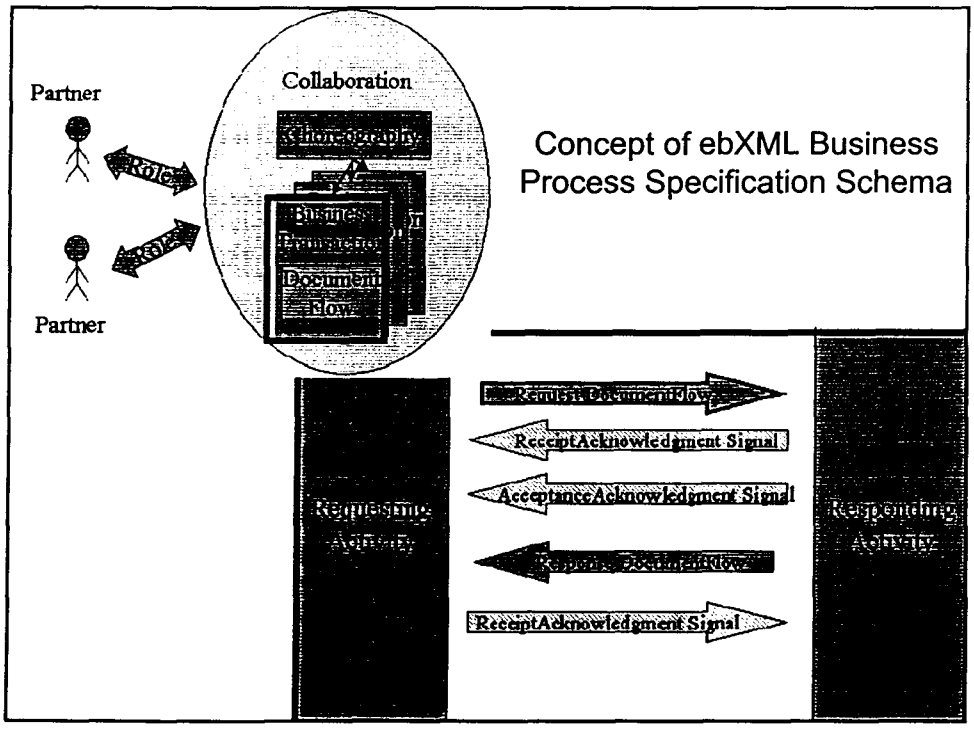


Process Modeling Ontology



Semantic Web Portal 35

Concept of ebXML Business Process Specification Schema



```
<BinaryCollaboration name="Firm Order"
timeToPerform="P2D">
     <Documentation>
          timeToPerform =
          Period: 2 days from start of transaction
     </Documentation>
     <InitiatingRole name="buyer"/>
     <RespondingRole name="seller"/>
     <BusinessTransactionActivity name="Create Order"
          businessTransaction="Create Order"
     fromAuthorizedRo:<BusinessTransaction name="Create Order">
          toAuthorize      <RequestingBusinessActivity name=""
</BinaryCollaboration>                isNonRepudiationRequired="true"
                              timeToAcknowledgeReceipt="P2D"
                              timeToAcknowledgeAcceptance="P3D">
                         <DocumentEnvelope
                              BusinessDocument="Purchase Order"/>
                        </RequestingBusinessActivity>
                        <RespondingBusinessActivity name=""
                              isNonRepudiationRequired="true"
                         timeToAcknowledgeReceipt="P5D">
                         <DocumentEnvelope isPositiveResponse="true"
                              BusinessDocument="PO
                         Acknowledgement"/>
                         </DocumentEnvelope>
                        </RespondingBusinessActivity>
                    </BusinessTransaction>
```

Business Process Specification in XML

KA2

An Ontology-Based Community Web Portal

KA2

- Knowledge Annotation Initiative of the Knowledge Acquisition Community
- The basic scenario
  - WWW documents of the KS community were annotated according to the schema of an ontology.
  - The annotations enable intelligent access to these documents and infer implicit knowledge from explicitly stated facts and rules from the ontology.

The KA2 Ontology

Person-ontology

Class hierarchy

```
Person
  Employee
    Academic-Staff
      Lecturer
      Researcher
    Administrative-Staff
      Secretary
      Technical-Staff
    Student
      Phd-Student
```

Relations

Address, Affiliation, Cooperates-With, Editor-Of,Email, First-Name, Has-Publication, Head-Of-Group, Head-Of-Project, Last-Name, Member-Of-Organization, Member-Of-Program-Committee, Member-Of-Research-Group, Middle-Initial, Organizer-Of-Chair-Of, Person-Name, Photo, Research-Interest, Secretary-Of, Studies-At, Supervises, Supervisor, Works-At-Project

Publication-ontology

Class hierarchy

```
On-Line-Publication
Publication
  Article
    Article-In-Book
    Conference-Paper
    Journal-Article
    Technical-Report
    Workshop-Paper
  Book
  Journal
    IEEE-Expert
    IJHCS
    Special-Issue
```

Relations

Abstract, Book-Editor, Conference-Proceedings-Title, Contains-Article-In-Book, Contains-Article-In-Journal, Describes-Project, First-Page, Has-Author, Has-Publisher, In-Book, In-Conference, In-Journal, In-Organization, In-Workshop, Journal-Editor, Journal-Number, Journal-Publisher, Journal-Year, Last-Page, On-Line-Version, ...

Accessing the Community Web Portal

- Query capability
  - In F-Logic mechanism
- Navigating capability
  - As the easy-to-use front-end of the query mechanism

Query Capabilities

For all publications of the researcher "Steffen Staab".

```
FORALL Pub <-
EXISTS ResID ResID:Researcher[NAME->> Steffen "Staab";
                              PUBLICATION ->> Pub].
```

Two researchers cooperate,
if a Researcher X works at a Project Proj and
if a Researcher Y works at the same Project Proj and
X is another person than Y

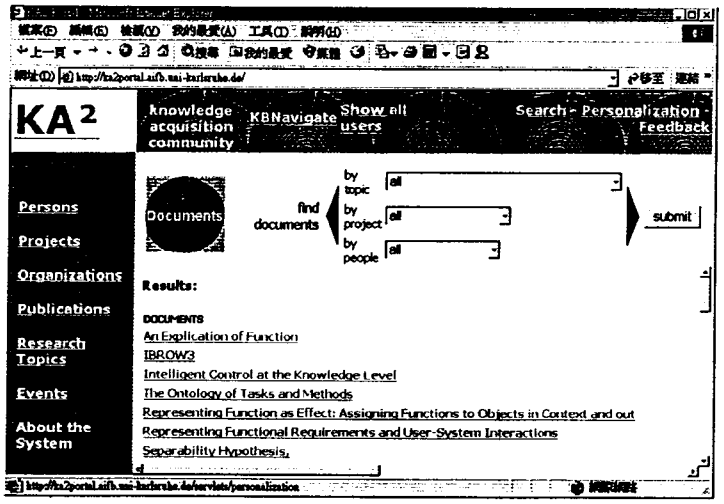
```
FORALL X; Y; Proj
 X:Researcher[COOPERATESWITH ->> Y:Researcher]
 <-
 X:Researcher[WORKSATPROJECT->>Proj:Project]
 AND
 Y:Researcher[WORKSATPROJECT->>Proj:Project]
 AND NOT equal(X; Y ):
```

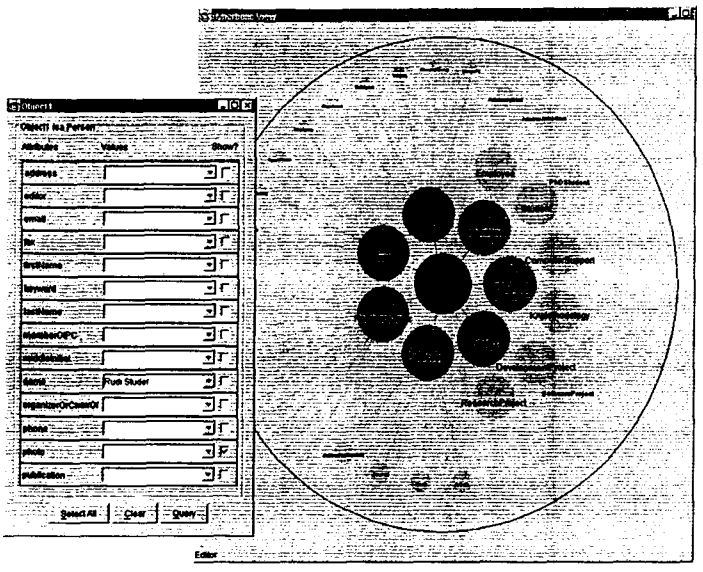
Which researchers are cooperating with other r?

```
FORALL ResID1, ResID2 <-
 ResID1:Researcher[COOPERATESWITH ->> ResID2].
```

Navigating and Querying the Portal

- Query the portal using F-Logic is too inconvenient.
- Navigating and querying the web
  - A *hypertext link* may contain a query dynamically evaluated when clicking on the link (See P. 45)
  - Query generated by using the *hyperbolic view interface* (See P. 46)
  - Queries may be personalized and for the different users and available for the user in a selection list.
  - An *expert mode*: Querying the portal by typing an F-Logic statement

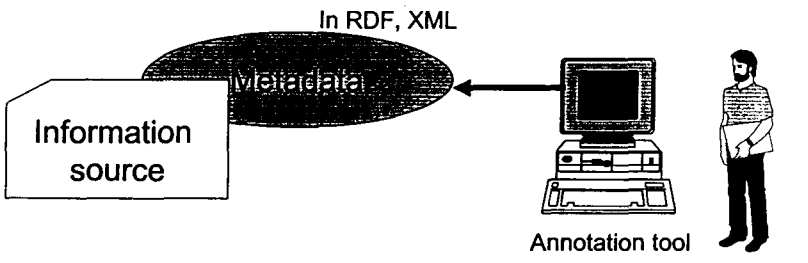




Providing Information

- Integrating various syntactic and semantic formats based on the common ontology
- Three different modes of information provision are supported
  - Metadata-based information
  - Wrapper-based information
  - Fact-based information

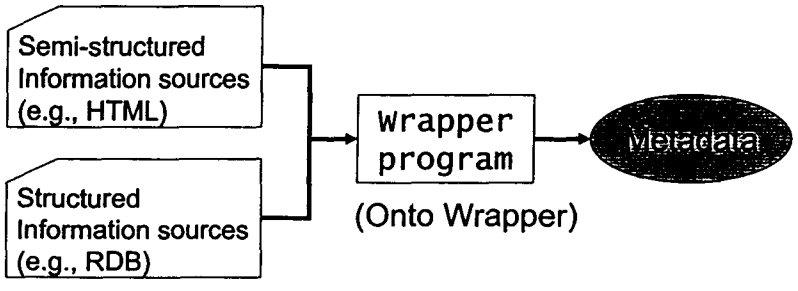
Metadata-based Information



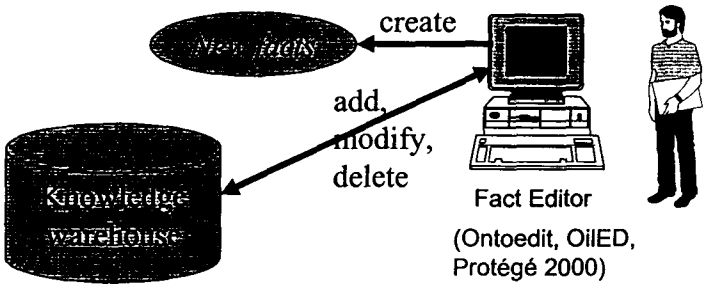
(OntoPad Annotea, ...)

Wrapper-based Information

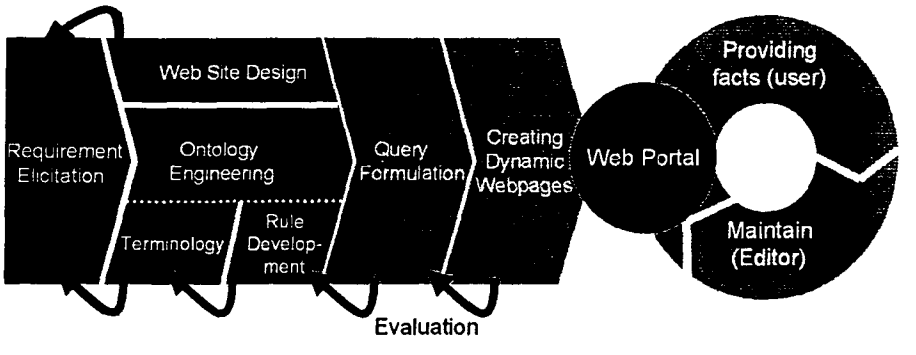
- Annotating information sources by hand is time-consuming.



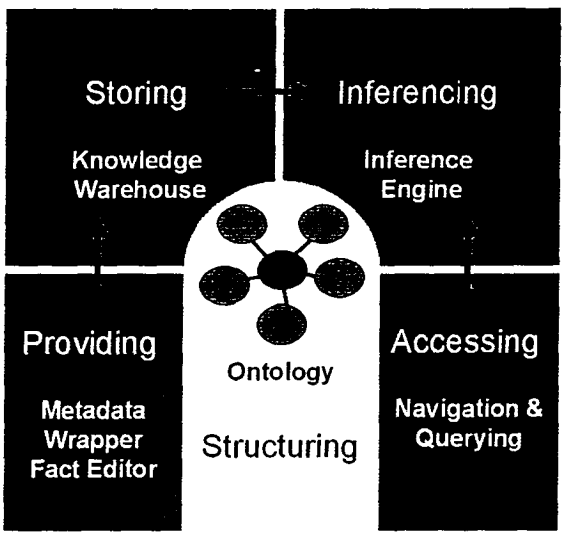
Fact-Based Information



Development of Web Portals

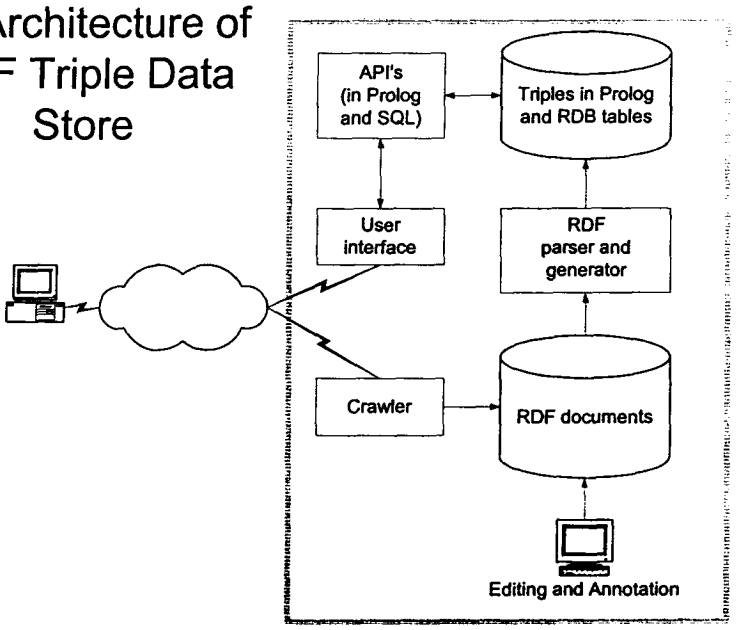


The System Architecture



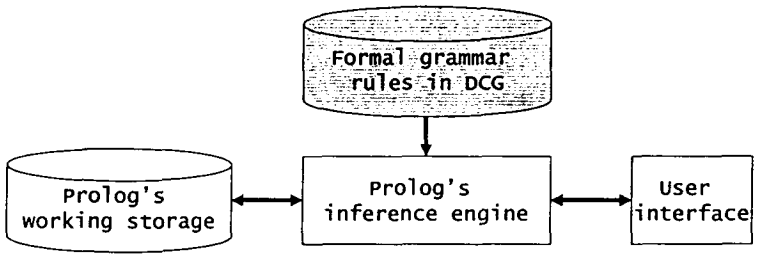
51

An Architecture of RDF Triple Data Store



RDF Parser and Generator

- Using DCG in Prolog as the development tool
- Formal grammar in RDF M&S Spec
  http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/#grammar




```
[6.1] RDF             ::= ['<rdf:RDF>'] obj* ['</rdf:RDF>']
[6.2] obj             ::= description | container
[6.3] description     ::= '<rdf:Description' idAboutAttr? bagIdAttr? propAttr* '/>'
                        | '<rdf:Description' idAboutAttr? bagIdAttr? propAttr* '>'
                          propertyElt* '</rdf:Description>'
                        | typedNode
[6.4] container       ::= sequence | bag | alternative
```

```
rdf(Objs) -->
   (['<?'],
    ([nm/'xml'];[nm/'XML'];
     [nm/'xmls'];[nm/'XMLS']),
    ([nm/'version','=',qs/_];[]),
    ([nm/'encoding','=',qs/_];[]),
    ['?>'],{wl('XML heading')}
     ;[]),
   (fullSTG('RDF',NS)
      ;
    halfSTG('RDF',NS),
       ['>']
      ;
    [ ]),
   objStar(Objs),
   (fullETG('RDF',NS);[ ]).

objStar(Out) -->
   obj(O),
      objStar(R),{Out=[O|R]}
      ;
   [],{Out=[]}.
```

```
obj(Obj) -->
   container(Obj)
      ;
description(_),{getAllTriples(Obj)}.
getAllTriples(Obj):-
findall(statement(A,B,C),statement(A,B
,C),Obj).

description(_) -->
   halfSTG('Description',NS),
   (idAboutAttr(IdAboutAttr);[]),
   (bagIdAttr(BagIdAttr);[]),
   propAttrStar(IdAboutAttr),
   (['/>']
      ;
     ['>'],
propertyEltStar(IdAboutAttr),
      fullETG('Description',NS)),
{reificationOfStatements(IdAboutAttr,B
agIdAttr)}
   ;
   typedNode.
```

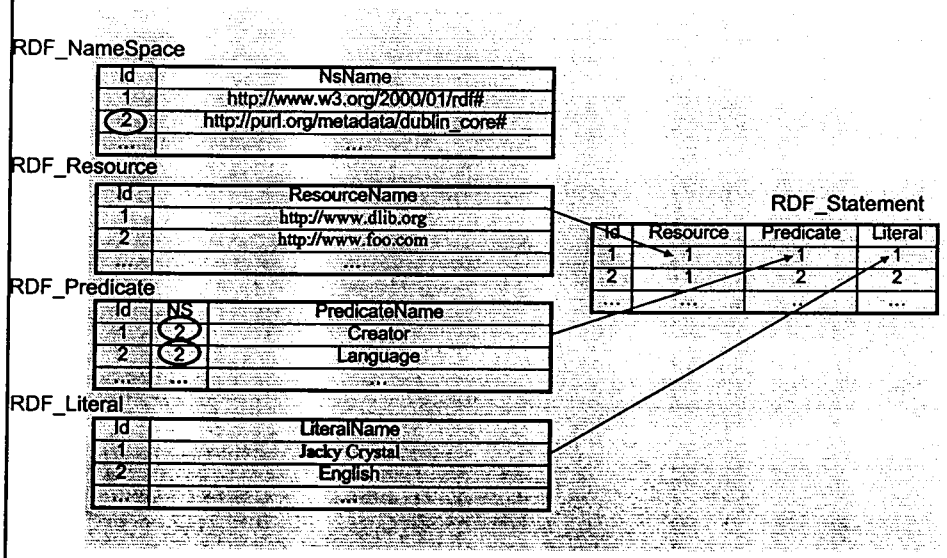
RDF input

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/metadata/dublin_core#">
  <rdf:Description about="http://www.foo.com/cool.html">
    <dc:Creator>
      <rdf:Seq ID="CreatorsAlphabeticalBySurname">
          <rdf:li>Mary Andrew</rdf:li>
          <rdf:li>Jacky Crystal</rdf:li>
      </rdf:Seq>
    </dc:Creator>
    <dc:Identifier>
      <rdf:Bag ID="MirroredSites">
          <rdf:li rdf:resource="http://www.foo.com.au/cool.html"/>
          <rdf:li rdf:resource="http://www.foo.com.it/cool.html"/>
      </rdf:Bag>
    </dc:Identifier>
    <dc:Title>
      <rdf:Alt>
          <rdf:li xml:lang="en">The Coolest Web Page</rdf:li>
          <rdf:li xml:lang="it">Il Pagio di Web Fuba</rdf:li>
      </rdf:Alt>
    </dc:Title></rdf:Description></rdf:RDF>
```

Prolog output

```
[statement(about/http://www.foo.com/cool.html,
          Title,
          alt(_87882,[string/The Coolest Web Page,
                      string/Il Pagio di Web Fuba])),
 statement(about/http://www.foo.com/cool.html,
          Identifier,
          bag(id/MirroredSites,[resource/http://www.foo.com.au/cool.html,
                                resource/http://www.foo.com.it/cool.html])),
 statement(about/http://www.foo.com/cool.html,
          Creator,
          seq(id/CreatorsAlphabeticalBySurname,[string/Mary Andrew,
                                                string/Jacky Crystal]))]
```

RDF Statements in RDB Tables



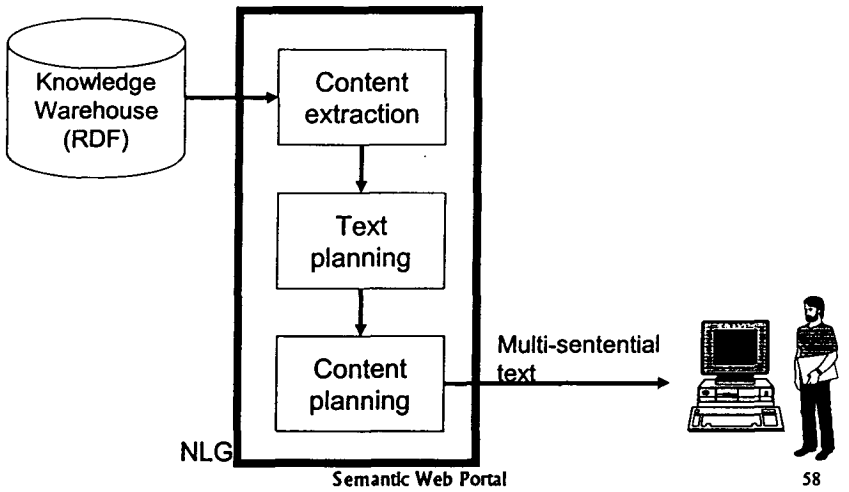
User Interface
Conceptual Search

- Natural language interface is the best choice, but is very difficult to obtain a good result.
- Form-based interface is easy to implement, but is too restricted.
- Hierarchical directory browsing is perhaps a good idea, if the domain in question is well organized.
- We also propose a frame-based query interface.

target_concept[(*attribute*$_1$[:*value*$_1$], …)]

`teacher(office:6th_floor)`

User Interface
Natural Language Generation



Future Work

- Building an ontology for the Digital Archive Services
  - Consisting of resource and process parts
- Declarative approach to semantic web service
  - Using an workflow engine to interpret the process descriptions
- Intelligent Q&A for digital archive

Summary

- Semantic Web portals
  - Machine-understandable information
  - RDF store
  - Accessing information
    - Navigation and query
  - Providing information
    - Annotation, wrapper, fact editing
  - Enabling automatic processing by software agents

Semantic Web and Agents

許永真
國立台灣大學資訊工程學系

Information Revolution

- Computers for scientific computation
- Computers as productivity tools
- Personal and home computers
  - Learning and research tools
  - Inter-person communication
  - Home entertainment
  - Information appliances
- Explosive growth of the Web
  - 1000 users in 1990.
  - Over 1 billion users device-wide by 2003.

Common Uses of the Web

- Information search and retrieval
- News and entertainment
- Shopping
- Travel: reservation/purchase/check-in
- Online community
- B2B applications
- P2P applications

A Futuristic Scenario

- During your daily workout, you heard a beautiful song playing over the Internet radio. Unfortunately, your cell phone signaled that you had a call from an important client. Before picking up the call, you commanded your PDA immediately to get a copy of the song for your personal collection so that you can enjoy it at a later time. Meanwhile, the volume of the radio turned down automatically.

Behind the Scene

- Your PDA will
  - Retrieve the current play list from the radio station.
  - Discover the song for purchase at several online stores.
  - Compare the price/feature and decide on a vendor.
  - Purchase the song on your behalf w/ your online wallet.
  - Schedule the song to be played later.
- Meanwhile,
  - Your cell phone will announce your picking up the call.
  - Your radio will adjust the volume based on pre-defined conditions on its volume control.

What's Missing?

- Taskable agents
  - Goals, believes, & plans
  - Making decisions and taking actions for their users.
- Machine-readable information sources
- Mechanism for utilizing unstructured, heterogeneous, & distributed information
- Communication among web-enabled devices
- Identification of trusted services/information

Web o Semantic Web

- Most of the Web's content today is designed for humans to read.
- Computers have no reliable way to manipulate & process the semantics.
- The Semantic Web will bring structure to the meaningful content of web pages so software agents can carry out sophisticated tasks for users.

The Semantic Web

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

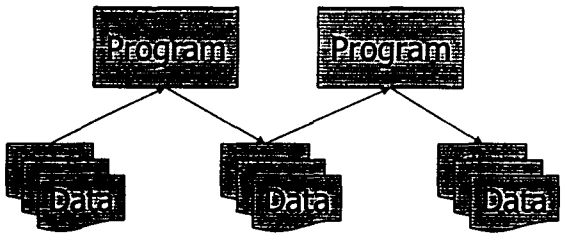
- *Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001*

Evolution of Information Processing

- Information becomes more
  - Complex
  - Structured
  - Meaningful
- Processing becomes increasingly
  - Distributed
  - Interactive
  - Intelligent

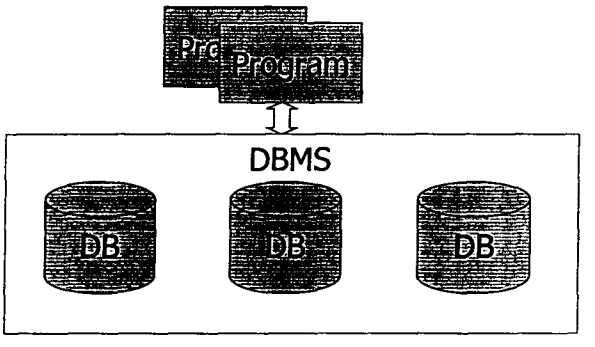
Stand-Alone Processing

- Program = Algorithm + Data
- Information is stored in data files.
- Programs may share information only by using data syntax defined for each data file.



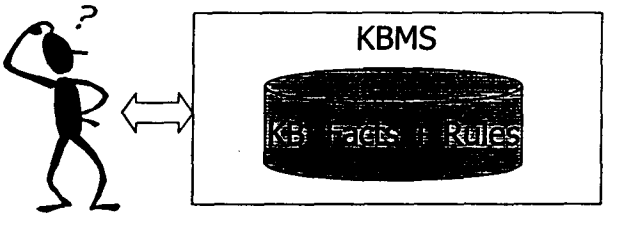
Database

- Information as structured data: records
- Data schema & query language



Knowledge Base

- Semantics: terms & relations
- Rules of inference
- Knowledge base query language

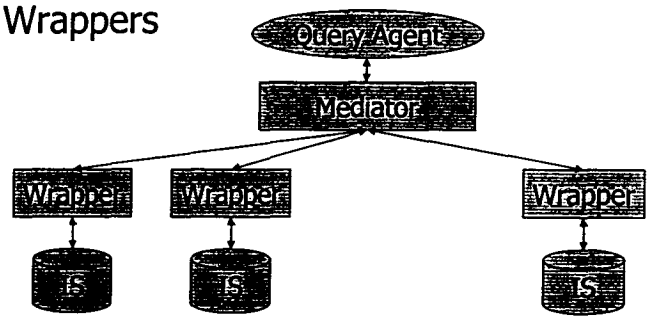


KR vs. Web

- Limited expressiveness anything goes!
- Brittle and demand consistency paradoxes and unanswered questions
- Facts non-sterilized information
- Centralized decentralized
- Domain-specific General
- Small scalable
- Declarative data Multimedia data

Querying the Web

- Search engines
- Web directories
- Information extraction w/o semantics
  - Wrappers



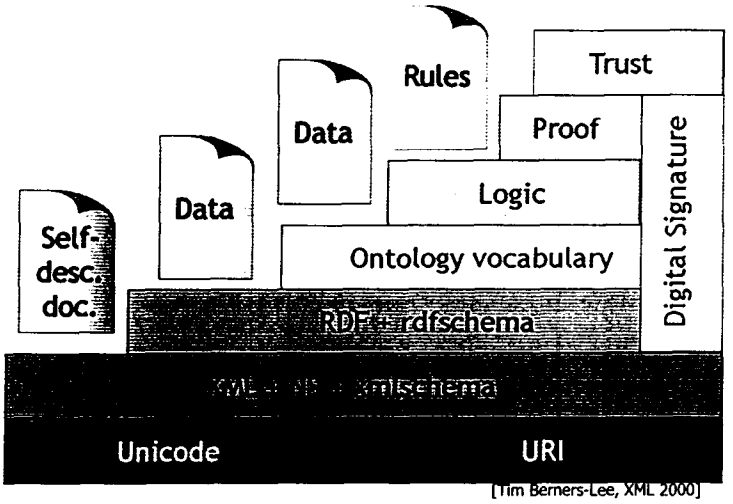
Web Knowledge Base

- Goal: To develop a probabilistic, symbolic knowledge base that mirrors the content of the world wide web. If successful, this will make text information on the web available in computer-understandable form, enabling much more sophisticated information retrieval and problem solving.
- Approach: developing a system that *can be trained* to extract symbolic knowledge from hypertext, using a variety of machine learning methods.
- [Mitchell et al., 1998]

Semantic Web (*Definition from W3C)*

- The **Semantic Web** is the abstract representation of **data** on the World Wide Web, based on the Resource Description Framework (RDF) standards and other standards to be defined.
- It is being developed by the W3C, in collaboration with a large number of researchers and industrial partners.

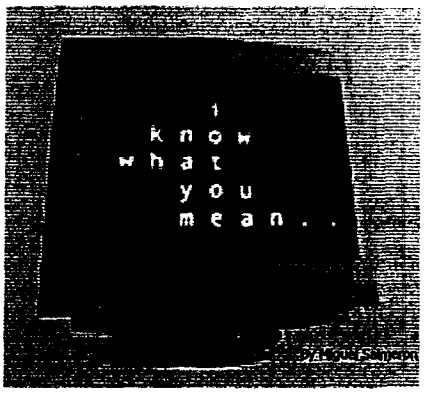
Semantic Web "Layer Cake"



[Tim Berners-Lee, XML 2000]

Semantic Web Agent

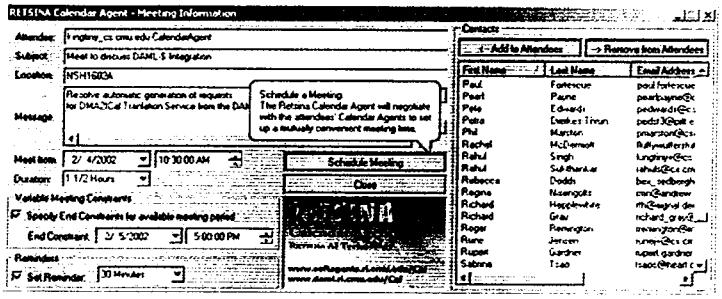
- Web content that is meaningful to computers
- Machine-readable ontology



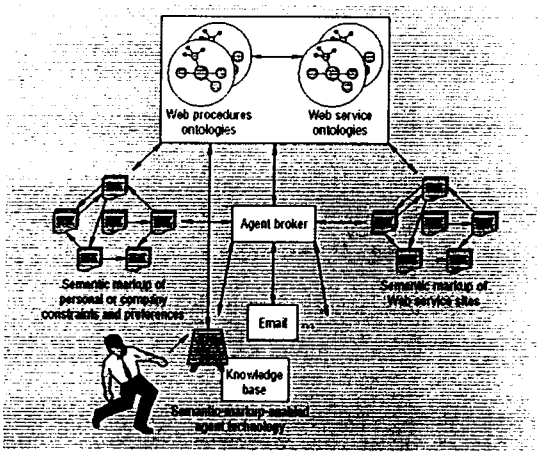
RETSINA Calendar Agent

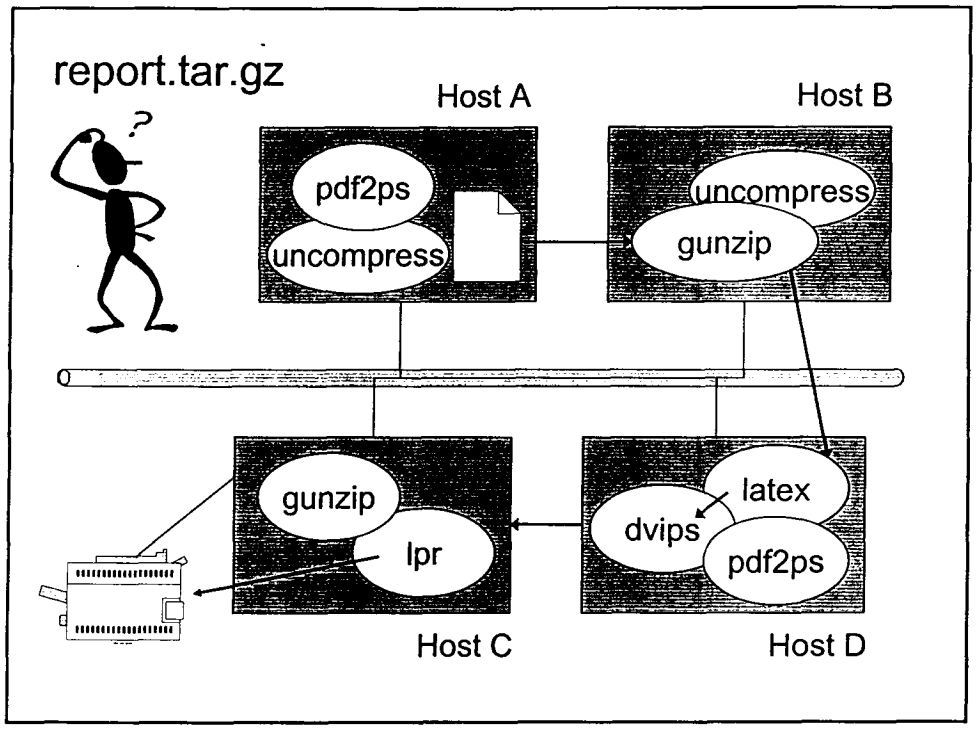
To schedule meetings between individuals based on their schedules maintained in MS Outlook

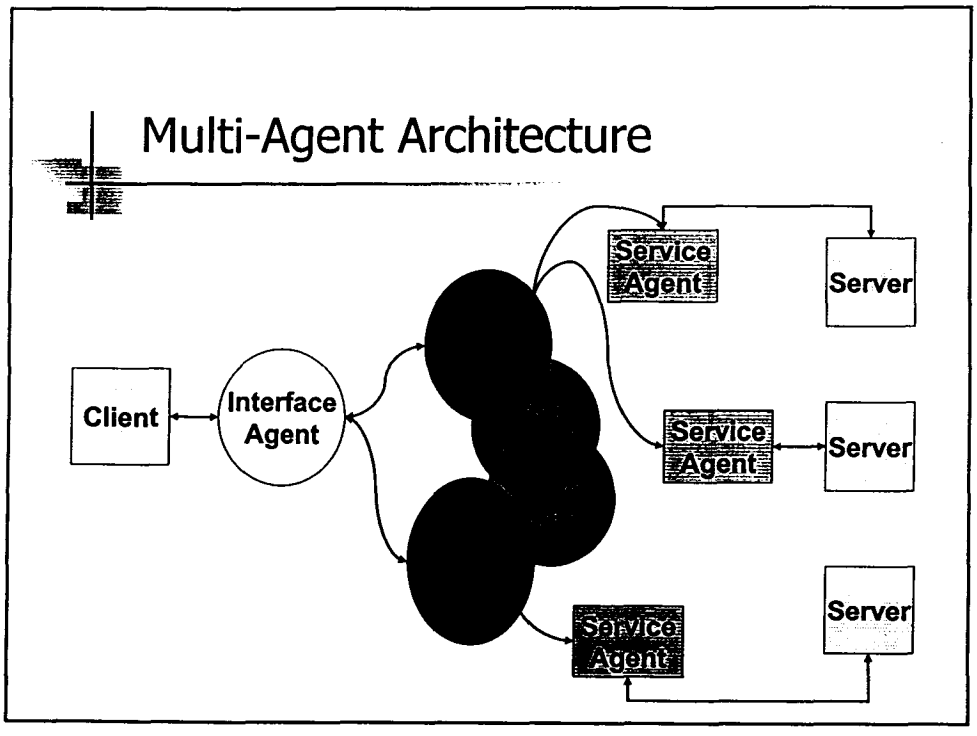
- Distributed Meeting Scheduling Engine
- RETSINA Semantic Web Calendar Parser



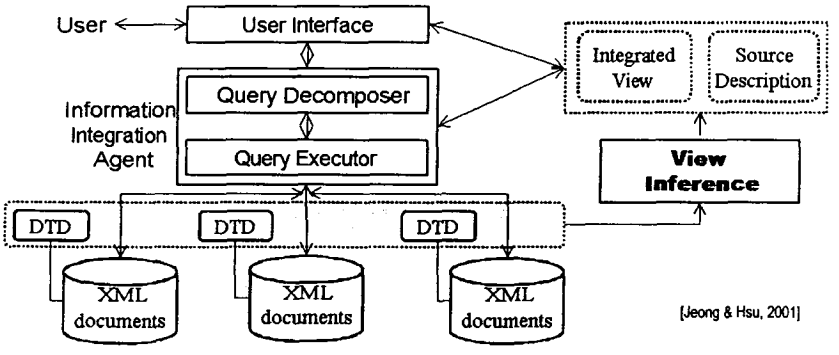
Semantic Web Services




report.tar.gz
Host A
Host B
pdf2ps
uncompress
uncompress
gunzip
gunzip
lpr
latex
dvips
pdf2ps
Host C
Host D

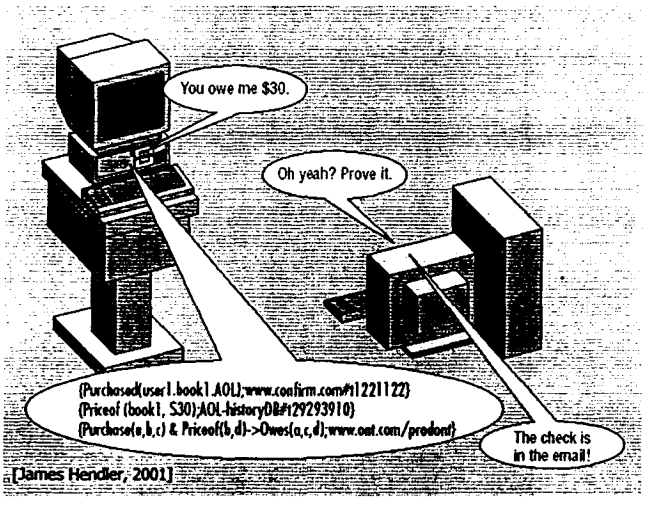

Multi-Agent Architecture
Service Agent
Server
Client
Interface Agent
Service Agent
Server
Service Agent
Server

XML Information Integration Agent



[Jeong & Hsu, 2001]

Trust and Proofs



[James Hendler, 2001]

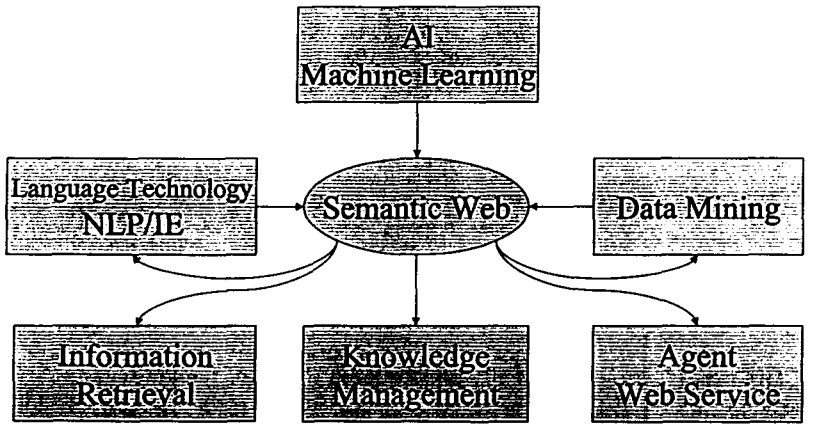
Summary

- The power of semantic web depends on successful integration of research on multiple disciplines:
  - Artificial intelligence
  - Information retrieval
  - Linguistics
  - Distributed systems & Web technology
- Semantic web and agents will empower machines to collaborate with humans more effectively.
- Agents will exploit users' constraints and preferences to help customize users' requests for automatic Web service discovery, execution, or composition and interoperation.

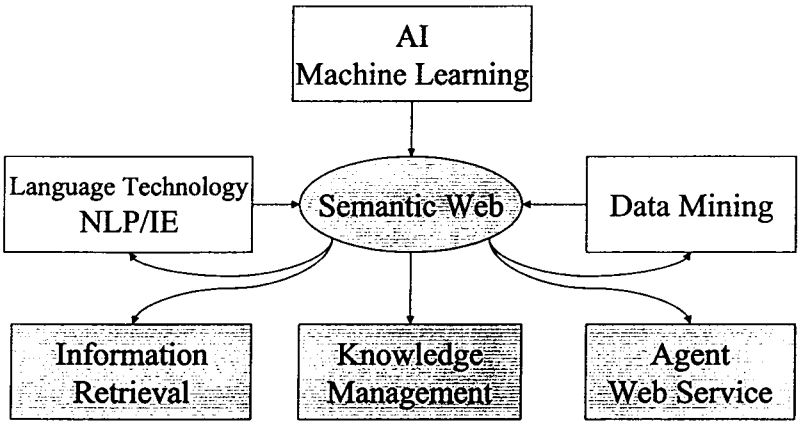
Applications of Semantic Web

Lin, Shih-Jui and Chien, Lee-Feng
Institute of Information Science
Academia Sinica

Semantic Web and Related Fields



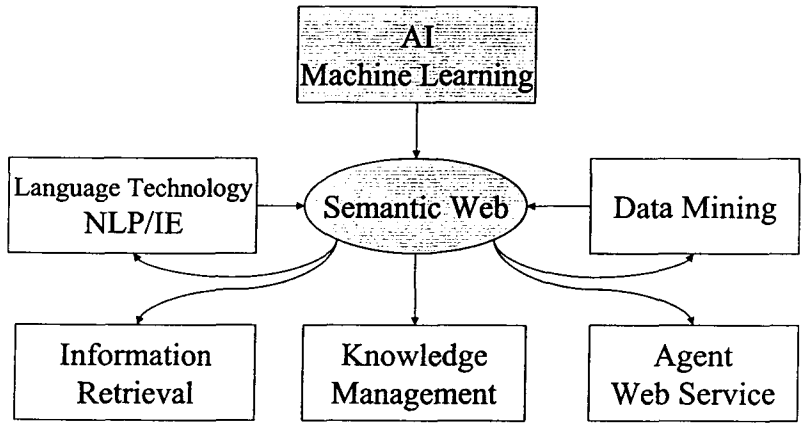
Semantic Web and Related Fields



Building Semantic Web

- Ontology
  - Building
    - repositories of terms and their relationships (LT)
    - ontology generation (ML)
  - Mapping and merging
    - knowledge of language, terms (LT)
    - mapping and merging (ML)
- Knowledge base
  - Adding instances into KB
    - structure/content mining (DM)
    - text analysis and extract values of attributes (NLP, IE, ML)
- Document
  - Semantic annotation
    - association between words and annotations (DM, ML)

Semantic Web and Related Fields



Using Semantic Web

- Language technology
  - Text corpora with semantics
- Data mining
  - Content/structure mining from semantic web pages
  - Usage mining from user's activities on semantic web

Using Semantic Web

- Information retrieval
  - Metadata search
  - Topic-based search
- Knowledge management
  - Acquire, maintain, access knowledge
- Agent technology / web services
  - DAML-S
  - RETSINA calendar agent

Application I

Information Retrieval

Information Retrieval

- Current search
- Search on Semantic Web
  - Metadata search
    - Project: HOWLIR
  - Topic-based search
    - Project: TAP

Current Search

- Is keyterm-based search (e.g., Google)
  - Full text indexing
  - Page authority (link analysis)
  - Page popularity (user's click)
- Problems
  - Not specific
    - Data in pages have no semantic annotations
    - Yo-yo Ma's most recent CD
  - No topic disambiguation
    - Documents with different topics mix together
    - Yo-yo Ma's CDs, concerts, biography, gossips…,

Information Extraction

- Wrapper
  - Specific web sites
  - Structured documents
  - Heuristic extraction
- Information extraction
  - Unstructured documents
  - Natural language analysis
  - Values for specific attributes
- Problems
  - Not flexible

    Current web provides little metadata
  - No topic disambiguation

XML

- Metadata

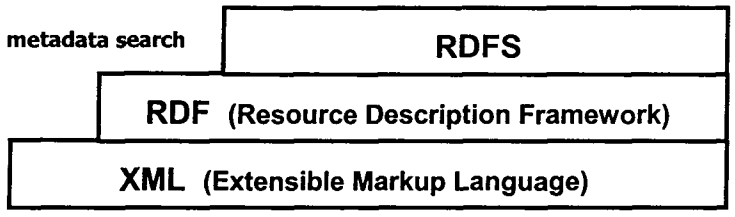
```
- <Person>
    <Name> Yo-yo Ma </Name>
    <CD>Inspired by Bach</CD>
  </Person>
```

XML (Extensible Markup Language)

Adapted from Dieter Fensel

RDF/RDFS

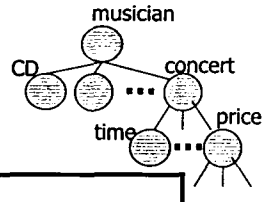
- Pre-defined modeling primitives
- The base of metadata search

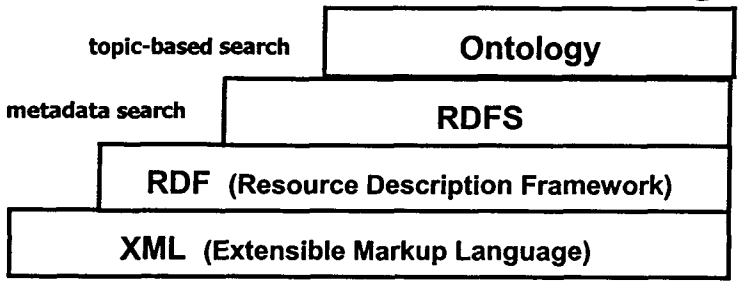


Adapted from Dieter Fensel

Ontology

- Sharable specifications of interesting topics
- The base of topic-based search





Adapted from Dieter Fensel

Search on Semantic Web

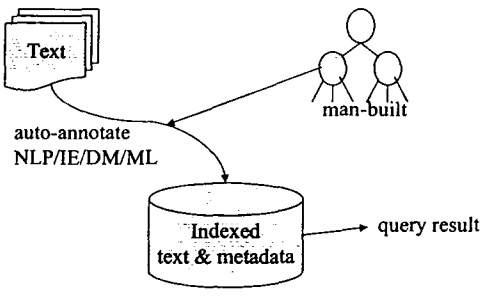
- Metadata search
  - To increase precision and flexibility
- Topic-based search
  - To help contextualize queries and overlay results in terms of a knowledge base

Metadata Search

- To annotate metadata on documents (XML/RDF/RDFS)
- To index both full text and metadata
- To retrieve documents according to both text and metadata (Hybrid IR)
- e.g., HOWLIR IR system (UMBC, John Hopkins)

HOWLIR

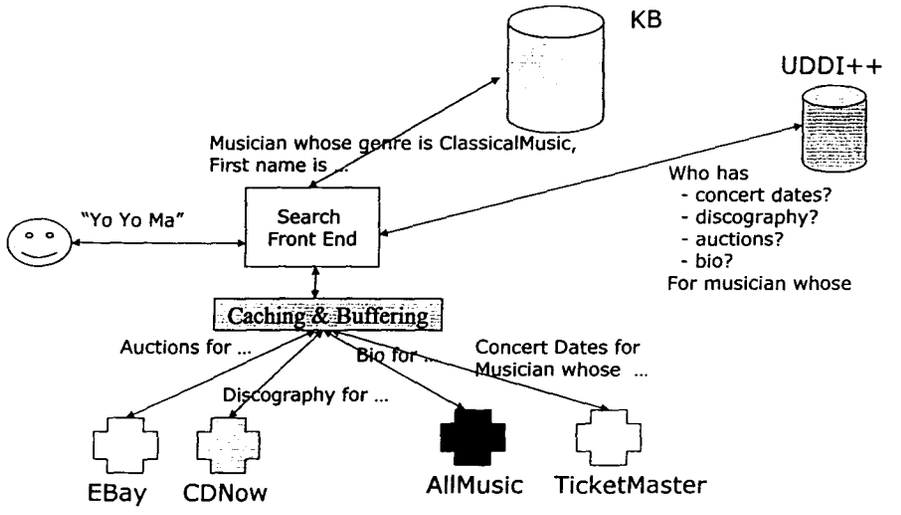
- To extract terms from documents via AutoText™
- To learn metadata by the statistical associations between metadata and text in annotated documents
- To generate annotations in RDF/DAML
- To retrieve documents according to text and metadata



Topic-based Search

- To help contextualize queries and overlay results in terms of a knowledge base
- E.g. TAP (IBM, Stanford)

TAP



TAP KB

- Ontology and instances in specific domains (music, sport, etc.)
  - Manual editing
  - Mining free data sources on the Web
  - Reading news articles and automatically identifying new musicians, athletes, etc.
- Currently covers about 20% of queries
- In RDF, DAML+OIL format
- Browse the KB at TAP site

Summary of IR

- Metadata search
  - HOWLIR
- Topic-based search
  - TAP

Application II

Knowledge Management

Knowledge Management

- What is KM?
- KM in a company
- KM on Semantic Web
- Project: Ontoknowledge

What is KM?

- Acquiring knowledge
  - Gather
  - Organize
- Maintaining knowledge
  - Represent
  - Update
- Accessing knowledge
  - Search
  - Visualize/browse
  - Share

KM in a Company

- To organize, maintain, and access the knowledge and experiences effectively (organization memory)
- To share documents among different departments
- To reduce the overhead of training
- To reduce the cost of customer services
- To reduce labor force

KM on Semantic Web

- Semantic web provides infrastructure for KM
  - Acquiring knowledge:
    - Ontology building
    - KB building
  - Maintaining knowledge:
    - Represented in RDF/DAML/OIL
  - Accessing knowledge:
    - Intelligent search
    - Ontology-based visualization
    - Ontology-based sharing

國科會 Web Technologies 系列研討會
Mobile Web Services Workshop



中華民國九十一年六月廿八日

國立成功大學國際會議廳

國科會 Web Technologies 規劃推動小組　　主辦

成功大學數位生活科技研究中心　　協辦

國科會 Web Technologies 系列研討會
Mobile Web Services Workshop

時　　間：中華民國 91 年 6 月 28 日（星期五）
地　　點：國立成功大學　光復校區　國際會議廳
主辦單位：國科會 Web Technologies 規劃推動小組
協辦單位：成功大學數位生活科技研究中心
說　　明：

結合行動通訊及 Web Services 是一個可以被預期的產業發展趨勢，本小組特委託成大數位生活科技研究中心規劃舉行 Mobile Web Services 研討會。

議　　程：

時間	內　　容	地點
8:30~10:30	論文發表	第三演講室
10:50~11:50	從３Ｇ看未來的行動理財服務 【富翊資訊　李明哲總經理】	第三演講室
11:50~13:00	Lunch	
13:00~14:00	座談 【主持人：成大電機系　蔡尚榮教授】	國際會議廳
14:10~15:10	無線網路服務的未來展望 【弘運科技　鄒伯川副總經理】	國際會議廳
15:10~15:20	Break	
15:20~16:20	行動式網際網路服務平台 【工研院電通所　鄭仁傑組長】	國際會議廳
16:20~16:30	Break	
16:30~17:30	為行動服務而設計之共通行動終端平台 【資策會嵌入式系統實驗室　王台中博士】	國際會議廳

