

<p>9.10</p> <p>Details of making recommendations</p> <ul style="list-style-type: none"> • Metadata and associated approaches • Pure full text based approaches • MAUT approaches • Content-based matching algorithms • Clique-based similarity approaches 	<p>9.12</p> <p>Widely used metadata standards</p> <ul style="list-style-type: none"> • Why standards? • Library sciences and digital libraries • Describe semantic content • Describe the technical details • Embedded <ul style="list-style-type: none"> eg meta tag within html • Separate <ul style="list-style-type: none"> Database of associated metadata • Independent <ul style="list-style-type: none"> eg. Rating service • Goals for standards <ul style="list-style-type: none"> – simplicity – interoperability – reliable semantics – flexibility 	<p>9.14</p> <p>Some standards</p> <ul style="list-style-type: none"> • Dublin Core - digital libraries <ul style="list-style-type: none"> http://dublincore.org/ (visited May 2002) • Multimedia - MPEG (Moving Picture Experts Group), Multimedia Content Description Interface <ul style="list-style-type: none"> http://mpeg.telecomitalia.com/standards/mpeg-7/mpeg-7.htm (visited May 2002) • EUNA - Educational Network, Australia for K-12, VET, Tertiary, adult, teacher resources and services <ul style="list-style-type: none"> http://www.edna.edu.au/ (visited May 2002) • GILS - Government Information Locator Service for US government documents <ul style="list-style-type: none"> • AGLS - Australian Government Information Locator System <ul style="list-style-type: none"> http://www.govonline.gov.au/projects/standards/agls.htm (visited May 2002) 	<p>9.16</p> <p>Dublin Core</p> <ul style="list-style-type: none"> • Sub-elements <ul style="list-style-type: none"> eg <ul style="list-style-type: none"> DC:Creator DC:Creator:Affiliation DC:Creator:Email eg: <ul style="list-style-type: none"> DC:Date DC:Date:Created DC:Date:Modified • Schemes <ul style="list-style-type: none"> eg: taxonomies and standards <p>ISO8601 http://www.pv.vinnu.no/~tsua/ISO8601.html (visited May 2002)</p> <p>DC:Date = "1st of December 2002"</p> <p>DC:Date = "(SCHEME=ISO8601) 2002-12-01"</p> <ul style="list-style-type: none"> • Languages <p>DC:Title = "EPS Resources"</p> <p>DC:Title = "(LANG=URD) EPS Resources"</p> <p>(ISO 639 Code for the Representation of the Names of Languages</p> <p>http://www.ansis-open.org/cover/is639a.html visited May 2002)</p>
<p>9.11</p> <p>Data models</p> <ul style="list-style-type: none"> • Type 1 Data: resources users are interested in <ul style="list-style-type: none"> eg: books, lesson plans, teaching strategies, curriculum ideas, web pages, music objects • Type 2 Data: structured data about Type 1 data <ul style="list-style-type: none"> eg catalogue entry data for name, author, metadata and indexes of websites • Type 3 data: also metadata, but only useful for collections <ul style="list-style-type: none"> eg: Dewey Decimal Classification System <p><i>Measuring up Metadata: Toward a Data Model for Distributed Resource Discovery</i>, Albert Ip, Mike Currie, Ian Morrison, Jon Mason http://www.dls.siu.com/metadata/DataModel.html</p>	<p>9.13</p> <p>Meta tags</p> <ul style="list-style-type: none"> • Agreed elements <ul style="list-style-type: none"> eg: <ul style="list-style-type: none"> <meta name = "Description" content = "EPS Resource Set"> <meta name = "Keywords" content = "Personalisation, recommender"> • Matching user interests? 	<p>9.15</p> <p>Dublin Core</p> <ul style="list-style-type: none"> • 1995 Dublin, Ohio, USA • Content <ul style="list-style-type: none"> – Title, Subject, Description, Source, Language, Relation, Coverage • Intellectual property <ul style="list-style-type: none"> – Creator, Publisher, Contributor, Rights • Instantiation <ul style="list-style-type: none"> – Date, Type, Format, Identifier 	<p>9.17</p> <p>Example of Ratings - PICS</p> <ul style="list-style-type: none"> • Platform for Internet Content Selection • Initial goal -- children's access to suitable material • Numeric ranges <ul style="list-style-type: none"> – low (0), medium (1), high (2) http://www.w3.org/PICS/ (visited May 2002) • Self rating <ul style="list-style-type: none"> eg from above site <pre><head> <META http-equiv="PICS-Label" content=" (PICS:1.1 "http://www.gd.org/v2.5" Labels on "1994.11.05T08:15-0500" until "1995.12.31T23:59-0000" for "http://w3.org/PICS/Overview.html" ratings (sds 0.5 density 0 colorhue 1)) </head></pre> <ul style="list-style-type: none"> • Independent bureau for ratings

<p>9.18</p> <p>P3P - Metadata for privacy/access</p> <ul style="list-style-type: none"> Platform for Privacy Preferences Project standardized set of multiple-choice questions which indicates how a site handles personal information about its users P3P enabled browsers "read" this and compare it to user's own set of privacy preferences. Location of site privacy policy Physical Contact Information Online Contact Information Unique Identifiers Financial Account Identifiers Computer Information Navigation and Click-stream Data Interaction Data Demographic and Socio-economic Data Preference Data Content <p>http://www.w3.org/P3P/ (visited May 2002)</p>	<p>9.20</p> <p>Pure full text based approaches</p> <ul style="list-style-type: none"> classic web Harvest - glimpse 'powerful indexing and query system that allows you to search through all your files very quickly' http://www.sclimunczides/Glimpse/ (visited May 2002) - eg 'unbelievable' with one typing error glimpse - 1 unbelievable - occurrences of "arizona" in all files with "mail" in their name glimpse - fr mail arizona - documents with both terms glimpse 'Arizona desert/windsurfing' Harvest project <p>C. Mc Bowman, Peter B. Danzig, Ueli Mamber and Michael F. Schwartz. <i>Scalable Internet Resource Discovery: Research Problems and Approaches</i>. Communications of the ACM, 37(8), pp. 98-107, August 1994.</p>	<p>9.21</p> <p>MAUT - SmartClient</p> <ul style="list-style-type: none"> user selects flights user can see relationship between price and flight duration for available flights user can set constraints during search eg departure time associated with airport chosen for start of trip <p>Pu, P. and Fallings, B. <i>Empowering buyers' experiences: the SmartClient approach</i>, in Proceedings of ACM CHI, 2000, pp 289-296.</p>	<p>9.20</p> <p>Retrieval by similarity</p> <ul style="list-style-type: none"> user specifies example OR user specifies set of attributes defines similarity metric <p>eg Entire restaurant recommender</p> <p>User selected aspects of importance: cuisine price style atmosphere occasion or user nominates an example of a restaurant</p> <p>Entire returns results and invites inputs:</p> <p>less \$\$, nicer cuisine (eg American, Cajun/Creole, Californian ... Asian, Chinese, Japanese ...), traditional - creative, livelier - quieter.</p> <p>Entire orders attribute importance</p> <ol style="list-style-type: none"> cuisine similarity to example given atmosphere similarity <p>User input + system prediction -> reduced search space</p> <p>eg cheaper \$\$ -> omit all more expensive</p> <p>Burke, R. Semantic Ratings and Heuristic Similarity for Collaborative Filtering. In AAAI Workshop on Knowledge-Based Electronic Markets, pages 14-20. AAAI, 2000.</p>
<p>9.19</p> <p>RDF - Resource Description Framework</p> <ul style="list-style-type: none"> Goal of interoperability lightweight ontology system -> semantic web Resource eg web page, its elements Property - a resource with a name and usable as a property eg title, author statement - combination of (resource, property, value) eg. (author, url1, Fred) XML allows creation of tags with text/data and other tags RDF allows definition of allowable syntax for them Metadata mappings <p>http://www.w3.org/RDF/ (visited May, 2002)</p>	<p>9.20</p> <p>MAUT approaches</p> <ul style="list-style-type: none"> Multi-Attribute Utility Theory (MAUT) Objects have a vector of attributes {a1, ..., an} Associated vector of importance weights {w1, ..., wn} Determine individual user's weighting vector Overall value = sum (a1w1 + ... + anwn) widely used 	<p>9.22</p> <p>MAUT - whose utility</p> <ul style="list-style-type: none"> Customer <ul style="list-style-type: none"> low cost interaction low price serves personal goals Seller etc <ul style="list-style-type: none"> higher profits (income, costs...) customers delivery of content encourage customer loyalty 	<p>9.21</p> <p>Cosine comparisons</p> <ul style="list-style-type: none"> Each object is represented as a vector indicating the presence or absence of each attribute $v1 = \{0, 1, 1, 0, \dots, 0, 1, 1\}$ $v2 = \{1, 0, 1, 0, \dots, 0, 1, 1\}$ Compare similarity of two objects $\text{cosine} = v1 \cdot v2 / (\text{sqrt}(v1) \cdot \text{sqrt}(v2))$ $= (0 \cdot 1 + 1 \cdot 0 + 1 \cdot 1 + 0 \cdot \dots) / (\text{sqrt}(0^2 + 1^2 + 1^2 + 0^2 + \dots) \cdot \text{sqrt}(1^2 + 0^2 + 1^2 + 0 \dots))$ Similar approaches use weighted model - weights are significance of each attribute {0, 10, 3, } {0, 8, 0, 1, }

Assessing effectiveness

Information Retrieval measures:

- Precision - fraction of recommendations correctly given
- Recall - fraction of correct recommendations found
- F1/ouf - Fraction of inappropriate recommendations

Recommender measures:

- Accuracy
- Learning Rate
- MAUT performance
- Affective