“This presentation is for informational purposes only and may not be incorporated into a contract or agreement.”
This document is for informational purposes. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described in this document remains at the sole discretion of Oracle. This document in any form, software or printed matter, contains proprietary information that is the exclusive property of Oracle. This document and information contained herein may not be disclosed, copied, reproduced or distributed to anyone outside Oracle without prior written consent of Oracle. This document is not part of your license agreement nor can it be incorporated into any contractual agreement with Oracle or its subsidiaries or affiliates.
THE INFORMATION COMPANY
Sandeepan Banerjee, Omar Alonso, Mark Drake, Meeten Bhavsar
Oracle Server Technologies

“This presentation is for informational purposes only and may not be incorporated into a contract or agreement.”
Toward an Information Grid
The Information Grid

- Imagine a web of data
- Readable by machines
  - Search, Aggregate, Relate, Transform, Report On, Mine Data – automatically
- Comprehensible by humans
  - Analyze and Visualize easily
- Scalable
  - Machines are cheap, and the price is dropping
  - Network is now faster than disk
- Flexible
  - Move data around without breaking the apps
Virtualization and Convergence

- Virtualization is a framework for dividing up the resources of an organization into multiple execution environments.
- Convergence seeks to bring together the management of all your data assets.
- The Information Grid combines the benefits of Virtualization and Convergence.
Evolution of the Information Grid
Infrastructure, Application and Information Grids

- **Infrastructure Grid**: Provisioning and failover
  - Virtualize coupling between processors, storage and running programs
  - Converge management

- **Application Grid**: SOA and Business Process Management
  - Virtualize coupling between client and server portions of a program
  - Converge delivery mechanisms and orchestrate

- **Information Grid**: Repository, Metadata, Search, Semantic Crawlers, Inferencing and Visualization
  - Virtualize coupling between data and applications
  - Converge seamless access to all kinds of data
Infrastructure Grid

- Enterprise Manager Grid Control
- Host management
- Performance tuning
- Remote administration
- Monitoring databases, application servers and applications
- Provisioning
- Extensible...
Application Grid

- Service-Oriented Architectures
  - IT infrastructure for request-reply applications
  - Application functions are modularized and presented as services
  - Services are loosely coupled
  - Interfaces are independent of implementation
Why SOA?
Business Process Management

• BPEL is emerging as the standard for assembling a set of discrete services into an end-to-end process flow
• Allows you to port business processes stored in one system to another by defining open standards-based interfaces
The Missing Layer

- The flexibility of a Grid cannot be realized
  - If application modules are too tightly coupled to data
    - If the creation or discovery of new kinds of data needs continuous UI or application code changes
  - If we cannot make more data machine readable
  - If we cannot provide seamless access to all kinds of data
  - If we cannot relate information across different sources, or analyze heterogeneous information
DEMONSTRATION

The Information Grid
Relating Information

• Search provides random access to data across sources
• Taxonomic classification provides dynamic categories which can be used to navigate better
• Ontologies help describe and relate information across sources
  • Better Decisions
What we will see

- Load the DBLP ontology and data into the Oracle XML DB Repository
  - Combine with Search, Clustering and Information Visualization for a rich, semantic search
  - Dynamic drill down into arbitrary structure across different sources and search for values, independent of location

- Focused semantic crawling to extract specific information from the web
  - Find pages on WWW that discuss DBLP researchers
  - Moving metadata on this information into XML DB

- Relating focused crawl information with other DBLP data
  - Inferencing of Home Pages
  - Building and Viewing Social Networks of Researchers
RDF

- A framework for describing and interchanging metadata
  - A **Resource** is anything that can have a URI – e.g. [http://www.oracle.com/technology/tech/xml/xquery/pdf/xquery10gr2v2.pdf](http://www.oracle.com/technology/tech/xml/xquery/pdf/xquery10gr2v2.pdf)
  - A **Property** is a Resource that has a name e.g. Author, Title, Type
  - A **Statement** consists of the combination of a Resource, a Property, and a Value

```xml
<rdf:Description about='http://www.oracle.com/technology/tech/xml/xquery/pdf/xquery10gr2v2.pdf'>
  <Title> XML Query (XQuery) Support in Oracle Database 10g Release 2 </Title>
  <Author>Sandeepan Banerjee</Author>
  <Home-Page rdf:resource='http://www.grandpoohbah.net/Sandeepan' />
</rdf:Description>
```
Ontologies

• An ontology is a means of capturing knowledge about a domain
  • Shared understanding
  • can be used both by humans and computers
• OWL is a Web Ontology Language
  • OWL builds on RDF and adds more vocabulary for describing properties and classes
  • relations between classes (e.g. disjointness)
  • cardinality (e.g. “exactly one”)
  • Equality
  • richer typing of properties
  • characteristics of properties (e.g. symmetry)
  • enumerated classes

<owl:Class rdf:ID="Masters Thesis">
<rdfs:subClassOf rdf:resource="#Document" />
</owl:Class>

<owl:Class rdf:ID="PhD Thesis">
<rdfs:subClassOf rdf:resource="#Document" />
<owl:disjointWith rdf:resource="#Masters Thesis" />
</owl:Class>

<owl:Class rdf:ID="Student Theses">
<owl:unionOf rdf:parseType="Collection">
<owl:Class rdf:about="#Masters Thesis" />
<owl:Class rdf:about="#PhD Thesis" />
</owl:unionOf>
</owl:Class>
DBLP

• The DBLP server provides bibliographic information on major computer science journals and conference proceedings

• Initially the server was focused on Data Base systems and Logic Programming, now it is gradually being expanded toward other fields of computer science, so as to become a Digital Bibliography & Library Project

• Started in the 1980s, now 660k articles (Aug 2005)

• http://dblp.uni-trier.de/
This XML file does not appear to have any style information associated with it. The document tree is shown below.

```xml
  - <owl:Ontology rdf:about=""/>
    <dc:title>DBLP Bibliography Collection</dc:title>
  - <dc:description>
    DBLP Bibliography Collection converted from its original XML format to RDF/OWL.
  </dc:description>
</owl:Ontology>
- <owl:Class rdf:ID="Mastersthesis">
  - <rdfs:subClassOf>
    <owl:Class rdf:ID="Document"/>
  </rdfs:subClassOf>
  <rdfs:label>Mastersthesis</rdfs:label>
  <rdfs:comment>A masters thesis.</rdfs:comment>
</owl:Class>
- <owl:Class rdf:ID="Collection">
  <rdfs:label>Collection</rdfs:label>
  - <rdfs:subClassOf>
    <owl:Class rdf:about="#Document"/>
  </rdfs:subClassOf>
  <rdfs:comment>A collection.</rdfs:comment>
</owl:Class>
- <owl:Class rdf:about="#Document">
  <rdfs:comment datatype="http://www.w3.org/2001/XMLSchema#string">A Document.</rdfs:comment>
</owl:Class>
```
Location Transparency

• Implementation of common protocols creates transparency between repositories
  • See files via a HTTP-server as Text/HTML
  • See a database as a filesystem via WebDAV
  • See an application as a RSS feed
  • …
• Move data between these repositories without breaking applications
processes lower systems queue types spaces networks solvers einem oriented computation model complexity invariants reasoning coordination inheritance estimation placement scheduling algebra task system vector polynomial linear much algorithm inductive machines logic constraint revisited external problem data libraries games success project sprache algebras processing calculus concept prozesse knowledge studies
<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tolga Yuruk</td>
<td>Efficient View Maintenance at Data Warehouses.</td>
<td>1997</td>
</tr>
<tr>
<td>Peter Van Roy</td>
<td>A Prolog Compiler for the PLM.</td>
<td>1984</td>
</tr>
<tr>
<td>Tatu Ylönen</td>
<td>Shadow Paging Is Feasible.</td>
<td>1994</td>
</tr>
<tr>
<td>Title</td>
<td>Author(s)</td>
<td>Year</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Leo Mark Self-Describing Database Systems - Formalization and Realization</td>
<td>(1985)</td>
<td></td>
</tr>
<tr>
<td>Peter Van Roy Can Logic Programming Execute as Fast as Imperative Programming?</td>
<td>(1990)</td>
<td></td>
</tr>
<tr>
<td>Gal Mitchell Extensible Query Processing in an Object-Oriented Database</td>
<td>(1993)</td>
<td></td>
</tr>
<tr>
<td>Lothar Breuer Spatial Queues</td>
<td>(2000)</td>
<td></td>
</tr>
<tr>
<td>Dmitry Efrosinan Controlled Queueing Systems with Heterogeneous Servers</td>
<td>(2004)</td>
<td></td>
</tr>
<tr>
<td>Joann J. Ordille Descriptive Name Services for Large Internets.</td>
<td>(1993)</td>
<td></td>
</tr>
<tr>
<td>Frank Olken Random Sampling from Databases</td>
<td>(1993)</td>
<td></td>
</tr>
<tr>
<td>Iakovos Motakis Temporal Reasoning in Active Databases.</td>
<td>(1997)</td>
<td></td>
</tr>
<tr>
<td>Michael H. Bohlen The Temporal Deductive Database System ChronoLog</td>
<td>(1994)</td>
<td></td>
</tr>
</tbody>
</table>
In the present thesis, a theoretical framework for the analysis of spatial queues is developed. Spatial queues are a generalization of the classical concept of queues as they provide the possibility of assigning properties to the users. These properties may influence the queueing process, but may also be of interest for themselves. As a field of application, mobile communication networks are modeled by spatial queues in order to demonstrate the advantage of including user properties into the queueing model. In this application, the property of main interest is the user's position in the network. After a short introduction, the second chapter contains an examination of the class of Markov-additive jump processes, including expressions for the transition probabilities and the expectation as well as laws of large numbers. Chapter 3 contains the definition and analysis of the central concept of spatial Markovian arrival processes (shortly: SMAPs) as a special case of Markov-additive jump processes, but also as a natural generalization from the well-known concept of BMAPs. In chapters 4 and 5, SMAPs serve as arrival streams for the analyzed periodic SMAP/M/c/c and SMAP/G/Infinity queues, respectively. These types of queues find application as models or planning tools for mobile communication networks. The analysis of these queues involves new methods such that even for the special cases of BMAP inputs (i.e. non-spatial queues) new results are obtained. In chapter 6, a procedure for statistical parameter estimation is proposed along with its numerical results. The thesis is concluded by an appendix which collects necessary results from the theories of Markov jump...
processes lower systems queue types spaces networks solvers einem oriented computation model complexity invariants reasoning coordination inheritance estimation placement scheduling algebra task system vector polynomial linear much algorithm inductive machines logic constraint revisited external problem data libraries games success project sprache algebras processing calculus concept prozesse knowledge studies
Fusion, Propagation, and Structuring in Belief Networks.

Stability Analysis of Wireless Networks

Ensembling neural networks: Many could be better than all.

Match Algorithms for Generalized Rete Networks.

Fusion and Propagation with Multiple Observations in Belief Networks.

Credal networks.

The Metaphorical Brain2: Neural Networks and Beyond (Michael A. Arbib).


Oscillating iteration paths in neural networks learning.

A Method for Isolated Thai Tone Recognition Using a Combination of Neural Networks.

Comparative Evaluation of Hypermesh and Multi-stage Interconnection Networks.

Microcomputer Networks.

A simple protocol for the dynamic tuning of the backoff mechanism in IEEE 802.11 networks.

Base station collaboration in Bluetooth voice networks.

On Reassembly Delay in Packet Switching Networks.

Granularity in all-optical WDM networks for large geographical areas.
Visualization

• A picture is worth a 1,000 words
  • In addition to pre-query refinement, enable post-query refinement through a visualization of results
  • Visual maps display what information is available in your documents, and how it is related and organized
• Quickly assimilate information
• Discovery
  • See patterns, leading causes, interconnections between documents, how documents are accessed.
Adding Compression to Block Addressing Inverted Indexes.

A dynamic storage allocation algorithm suitable for file structures.

Ricardo A. Baeza-Yates *Hierarchies of Indices for Text Searching.*

Improved Bounds for the Expected Behaviour of AVL Trees.

A Trivial Algorithm Whose Analysis is Not: A Continuation.

An Algorithm for String Matching with a Sequence of don't Cares.

An Analysis of the Karp-Rabin String Matching Algorithm.

Some Average Measures in m-ary Search Trees.

Height Balance Distribution of Search Trees.

Gonzalo Navarro *Very Fast and Simple Approximate String Matching.*

Ricardo A. Baeza-Yates *Optimal bounded disorder.*

Ricardo A. Baeza-Yates *A Framework to Animate String Algorithms.*

Ricardo A. Baeza-Yates *Fast and Practical Approximate String Matching.*

Fast Two-Dimensional Pattern Matching

Characterization of a protein binding sequence in the promoter region of the 16S rRNA gene of the spinach chloroplast genome.

Analysis of Linear Hashing Revisited.
Hierarchies of indices for text searching*1

Ricardo Baeza-Yatesa, Eduardo F. Barbosab and Nivio Zivianb

a Departamento de Ciencias de la Computación Universidad de Chile, Santiago, Chile
b Departamento de Ciência da Computação Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Received 17 August 1994; revised 2 August 1996. Available online 11 June 1999.

Abstract

We present an efficient implementation of a recently known index for text databases, when the database is stored on secondary storage devices such as magnetic or optical disks. The implementation is built on top of a new and simple index for texts called pat array (or suffix array).
AJAX

- Asynchronous Javascript And XML is a development technique for creating interactive web applications using a combination of:
  - HTML (or XHTML) and CSS for presenting information
  - The Document Object Model manipulated through Javascript to dynamically display and interact with the information presented
  - The XMLHttpRequest object to exchange data asynchronously with the web server
<table>
<thead>
<tr>
<th>Title</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling Variant User Requirements in Domain Engineering for Reuse.</td>
<td>note</td>
</tr>
<tr>
<td>Generation of Test Search Applications for Databases. An Exercise on Doma</td>
<td>note</td>
</tr>
<tr>
<td>Experiences in process domain engineering at PRC Inc.</td>
<td>note</td>
</tr>
<tr>
<td>Domain Engineering: a &quot;Radical Innovation&quot; for Software and Systems Engineering.</td>
<td>title</td>
</tr>
<tr>
<td>Patterns, Teams and Domain Engineering.</td>
<td>title</td>
</tr>
<tr>
<td>Using ontologies for the specification of domain-specific languages in multi-age environments.</td>
<td>title</td>
</tr>
<tr>
<td>Domain Engineering. A Software Engineering Discipline in Need of Research.</td>
<td>title</td>
</tr>
<tr>
<td>Jargons for domain engineering.</td>
<td>title</td>
</tr>
<tr>
<td>DL-based Support to Domain Engineering.</td>
<td>title</td>
</tr>
<tr>
<td>Domain Engineering: The Challenge, Status, and Trends.</td>
<td>title</td>
</tr>
<tr>
<td>An ontological approach to domain engineering.</td>
<td>title</td>
</tr>
<tr>
<td>Objects and Domain Engineering - Panel Session.</td>
<td>title</td>
</tr>
<tr>
<td>Software Asset Management and Domain Engineering.</td>
<td>title</td>
</tr>
<tr>
<td>Measuring Domain Engineering Effects on Software Change Cost.</td>
<td>title</td>
</tr>
<tr>
<td>Domain Engineering And Reuse.</td>
<td>title</td>
</tr>
</tbody>
</table>
Modeling Variant User Requirements in Domain Engineering for Reuse.

Generation of Text Search Applications for Databases. An Exercise on Domain Engineering.

Experiences in process domain engineering at PRC Inc.


Patterns, Teams and Domain Engineering.

Using ontologies for the specification of domain-specific languages in multi-agent domain engineering.

Domain Engineering: A Software Engineering Discipline in Need of Research.

Jargons for domain engineering.

DL-based Support to Domain Engineering.


An ontological approach to domain engineering.

Objects and Domain Engineering - Panel Session.

Software Asset Management and Domain Engineering.

Measuring Domain Engineering Effects on Software Change Cost.

Domain Engineering And Reuse.
What we will see

• Load the DBLP ontology and data into the Oracle XML DB Repository
  • Combine with Search, Clustering and Information Visualization for a rich, semantic search
  • Dynamic drill down into arbitrary structure across different sources and search for values, independent of location

  Focused semantic crawling to extract specific information from the web
  • Find pages on WWW that discuss DBLP researchers
  • Moving metadata on this information into XML DB

• Relating focused crawl information with other DBLP data
  • Inferencing of Home Pages
  • Building and Viewing Social Networks of Researchers
Inference-Based Retrieval

- The content of Web resources is in most part opaque to computers
  - Browsers display them and search engines locate words within them, but the level of "understanding" of the content is limited.
- A search engine, for example, might know that a resource contained the textual string “Alonso” but not that it was a representation of a Person, and that some Persons have home-pages, where a home-page is an URL that returns HTML.
- By enabling richer representation such as this, RDF makes it possible to express queries that go beyond simple text-matching.
Focused Crawler Configuration
This XML file does not appear to have any style information associated with it. The document tree is shown below.

```xml
<Homepages xmlns:location="Homepages.xsd">
  <hp>
    <name>Ricardo Baeza</name>
    <affiliation>University of Chile</affiliation>
    <url>http://www.dcc.uchile.cl/~rbaeza</url>
  </hp>
</Homepages>
```
<table>
<thead>
<tr>
<th>Browse ontology</th>
<th>Browse cluster - viz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article</td>
<td>processes</td>
</tr>
<tr>
<td>Book</td>
<td>lower system</td>
</tr>
<tr>
<td>Collection</td>
<td>queue types</td>
</tr>
<tr>
<td>Inproceedings</td>
<td>spaces</td>
</tr>
<tr>
<td>Masterthesis</td>
<td>networks</td>
</tr>
<tr>
<td>Phdthesis</td>
<td>solvers einer oriented</td>
</tr>
<tr>
<td>Proceedings</td>
<td>untersuchungse</td>
</tr>
<tr>
<td>Series</td>
<td>computation</td>
</tr>
<tr>
<td>Www</td>
<td>model</td>
</tr>
<tr>
<td></td>
<td>complexity</td>
</tr>
<tr>
<td></td>
<td>invariants</td>
</tr>
<tr>
<td></td>
<td>reasoning</td>
</tr>
<tr>
<td></td>
<td>coordination</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Browse ontology</th>
<th>Browse cluster - viz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive Optimization in a Database Programming Language</td>
<td>processes</td>
</tr>
<tr>
<td>Optimization of Large Join Queries</td>
<td>lower system</td>
</tr>
<tr>
<td>Nonclausal Logic Programming</td>
<td>queue types</td>
</tr>
<tr>
<td>Key Objects in Garbage Collection</td>
<td>spaces</td>
</tr>
<tr>
<td>The Design and Implementation of the SELF Compiler, an Optimizing Compile for Object-Oriented Programming Languages</td>
<td>networks</td>
</tr>
<tr>
<td>Glue: A Deductive Database Programming Language</td>
<td>solvers einer oriented</td>
</tr>
<tr>
<td>Subgoal Order for Query Optimization in Logic Databases</td>
<td>untersuchungse</td>
</tr>
<tr>
<td>Query Optimization in Deductive and Relational Databases</td>
<td>computation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Browse ontology</th>
<th>Browse cluster - viz</th>
</tr>
</thead>
<tbody>
<tr>
<td>university</td>
<td>affiliation</td>
</tr>
<tr>
<td>Source</td>
<td>Home pages</td>
</tr>
<tr>
<td>Attributes</td>
<td>url</td>
</tr>
<tr>
<td></td>
<td>hp</td>
</tr>
<tr>
<td></td>
<td>name</td>
</tr>
<tr>
<td></td>
<td>Homepages</td>
</tr>
</tbody>
</table>
What we will see

• Load the DBLP ontology and data into the Oracle XML DB Repository
  • Combine with Search, Clustering and Information Visualization for a rich, semantic search
  • Dynamic drill down into arbitrary structure across different sources and search for values, independent of location

• Focused semantic crawling to extract specific information from the web
  • Find pages on WWW that discuss DBLP researchers
  • Moving metadata on this information into XML DB

• Relating focused crawl information with other DBLP data
  • Inferencing of Home Pages
  • Building and Viewing Social Networks of Researchers
QUESTIONS & ANSWERS