



# Data Recognition Technology Technical Presentation





## General Features

- Search results are provided in sort order, by relevance
- Can be used to search on both structured and unstructured data
- Any subset of search fields may be used in any search
- Will isolate string fragments
- Can search any table or view of interest
- Will work on any relational database



## Search Properties for Managing Relevance

Home Information

Words:  Saved Screen:  Search Depth:  Weights

Phrase:  Results Cut-off:  Text Search:  Previous Next

Clear Export Search

Inexact	LastName	<input type="text"/>
Inexact	MiddleName	<input type="text"/>
Inexact	FirstName	<input type="text"/>
Inexact	Zip	<input type="text"/>
Inexact	DateofBirth	<input type="text"/>
Inexact	Address	<input type="text"/>

The desired query criteria can be specified for each field determined to be a Search Field. By default the fields search mode are set to InExact which will enable fuzzy processing. Additionally, you may specify Exact, Wildcard, OR, or Not as the mode for each field.

Details  
Occupation  
Employer

Search Object

Search Depth

Results Cut-off

Text Depth

Search Relevance (Weights)



## Open Search

Home Information

Words:  Screen: 1\_Person220Mil Search Depth: Standard Weights

Phrase:  Results Cut-off: Very Loose Text Search: Small

Clear Export Search Previous Next

Inexact LastName

Inexact MiddleName

Inexact FirstName

Inexact Zip

Inexact DateofBirth

Inexact Address

The desired query criteria can be specified for each field determined to be a Search Field. By default the fields search mode are set to InExact which will enable fuzzy processing. Additionally, you may specify Exact, Wildcard, OR, or Not as the mode for each field.

Details  
Occupation  
Employer

This type of search does not require any knowledge of fields to be searched

Words - will search all structured fields

Phrase - will search all unstructured fields

Can be done in any combination with structured search



# ALTER

## Search Booleans

Inexact (Default)

Exact

Or

Not

Wildcard



Use as an API

- One method: 'Search'
- Properties for:
  - Booleans
  - Parameters
  - Open Search
- Application linking
  - Identify where search arguments will be found
  - Identify where search results to be sent



## Search Object Generator

Creates the API's search objects

Administrator identifies the DB,  
and then the table or view

For each field selected, the  
administrator identifies if the  
field is for search purposes or  
for display purposes (any field  
can be both)

For a search field, the  
administrator also identifies  
the search type

**Connections and Screens:**

Database: OVERMIND\_REBOF [New] [Delete]  
Screen: 1\_Person220MII\_Act [New] [Save] [Delete]  
Database Address: 100.0.0.3\SQL2008  
Port: 1433  
Login: cylab

**Screen Info:**

**Tables:**  
MiddleNames\_Demo  
MiddleNames\_Demo\_Munged  
MultipleMatches  
MultipleSourceMatches  
MultipleTargetMatches  
Occupations  
Otrack\_Report  
Person

**Fields:**  
Address  
Comment  
DateOfBirth  
Employer  
FirstName  
LastName  
MiddleName  
Occupation

**Roles:**  
Administration User

**Search Fields:**

Type	Field	Display Name
Standard	LastName	LastName
Standard	MiddleName	MiddleName
Standard	FirstName	FirstName
Comment	Address	Address
Standard	Zip	Zip
Standard	DateOfBirth	DateOfBirth

**Display Fields:**

Field	Display Name
FirstName	FirstName
MiddleName	MiddleName
LastName	LastName
Address	Address
Zip	Zip
DateOfBirth	DateOfBirth

**Detail Fields:**

Field	Display Name
Occupation	Occupation
Employer	Employer
Address	Address



## Search Objects

**Search Objects** are the class of objects used by Vizier™ to define:

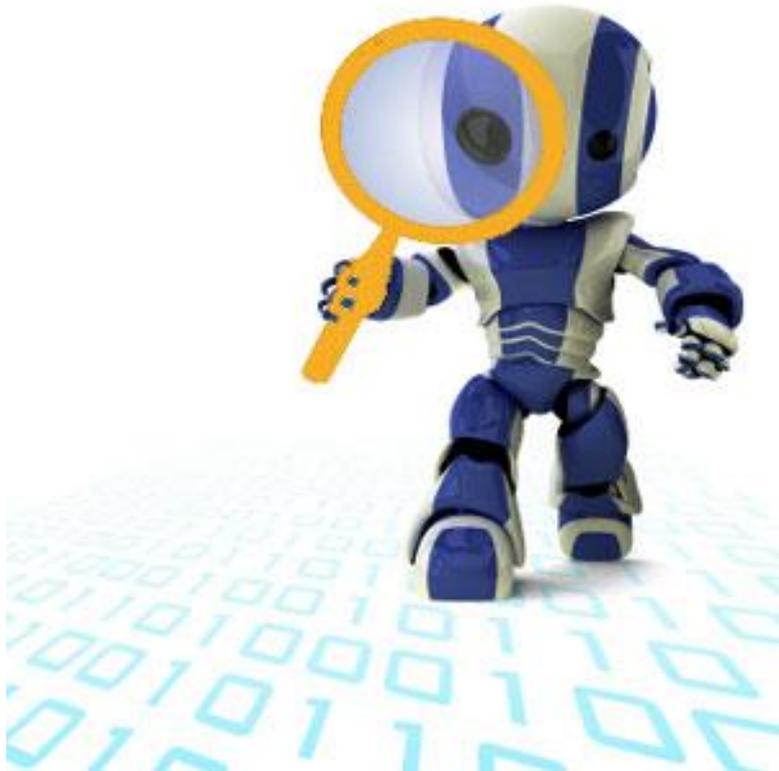
What fields the user can search on

How those fields are to be searched

Structured

Unstructured

Where to send the search results



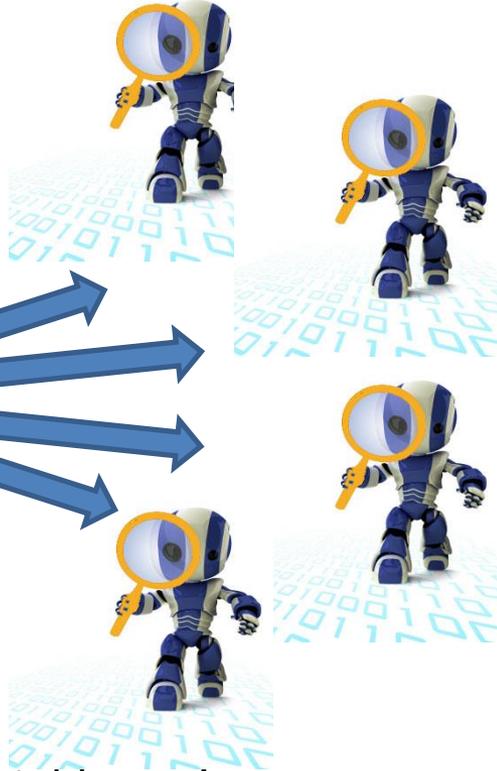
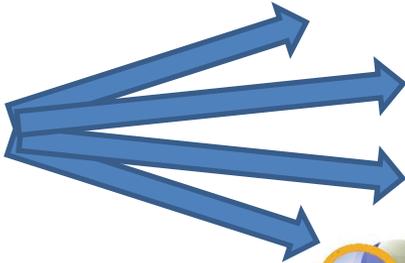


# Search Object Generator

Once the search object is tied to a table or view in a database...



**Vizier™**  
**Administrator**



An unlimited number of “search objects” can be created for the table or view  
Each “search object” can be used by an unlimited number of users



## The Vizier™ Solution

The Vizier™ solution is designed to support application recognition of data as an API. It is intended as an **enterprise level solution** utilizing data recognition **as a service**.



Searching



Matching



Preventing  
Duplication



## Enhance Save and Search Functionality

Search: 1760781766248 any of these As: ISBN (no hyphens/spaces)

AND OR NOT

Search: any of these As: Title

AND OR NOT

Search: any of these As: Title

25 items displayed per page

Search

Search

Customers

Summary

Customer# ANTON

Company Name Antonio Moreno Taqueria

Contact Name Antonio Moreno

Contact Title Owner

Address Mataderos 2312

City México D.F.

Region

Postal Code 05023

Country Mexico

Phone (5) 555-3932

Fax

Create

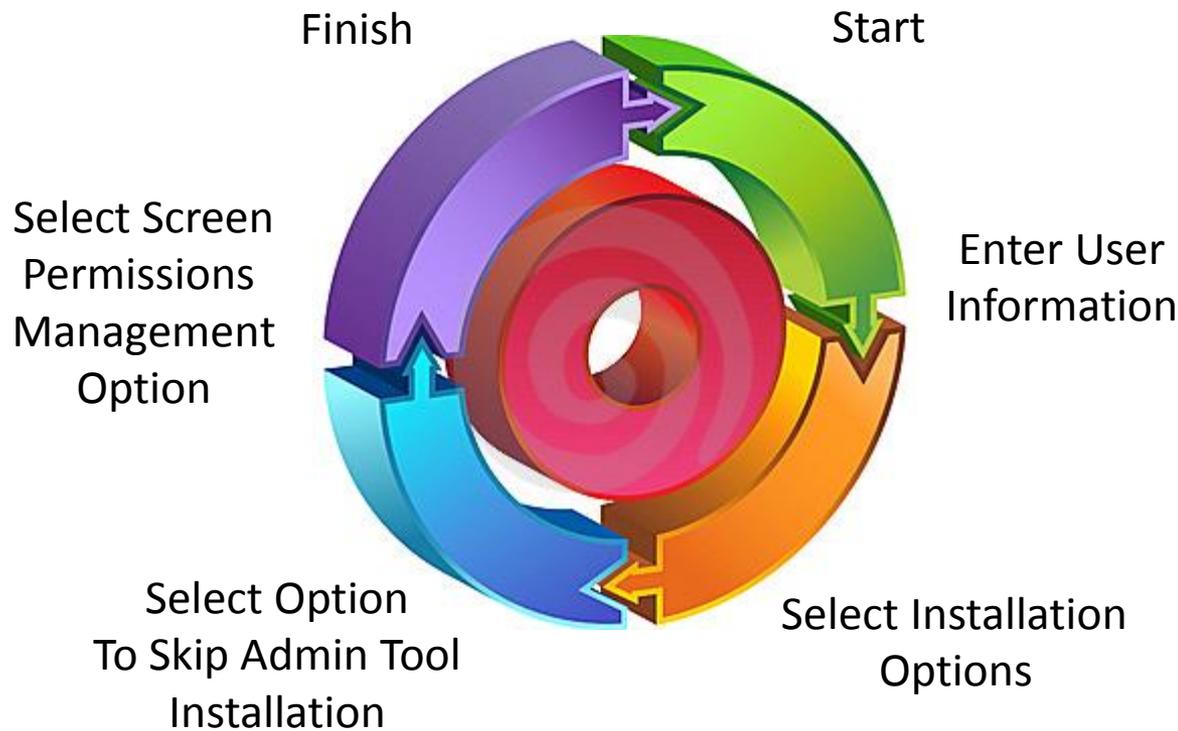
The API can be included as part of a procedure behind special purpose buttons that are intended to exploit data recognition capabilities

- Enable fuzzy searching on search or lookup screens
- Prevent record duplication on save screens



# Installation Wizard

Can be used to install:  
Standalone Software  
Administration Tool





## Documentation



- Release Notes
- Installer's Guide
- User's Guide
- Administrator's Guide



**Cylab Inc.**

8500 Leesburg Pike

Suite 405

Vienna, VA 22182

Office: (703) 379-4818

Fax: (703) 379 – 1336

E-mail: [lorenzow@cylabinc.com](mailto:lorenzow@cylabinc.com)



## What Types of Data Problems Does DRT Help to Overcome?



Special Characters

Format Details

Spacing,  
Punctuation  
Dates

Text Component Reversals

Extraneous Keystrokes

Misspellings

And Much More!



The Vizier™ solution is independent of:



Subject or Topic

Database Vendor

User's Language

Data Type

Data Model Structure



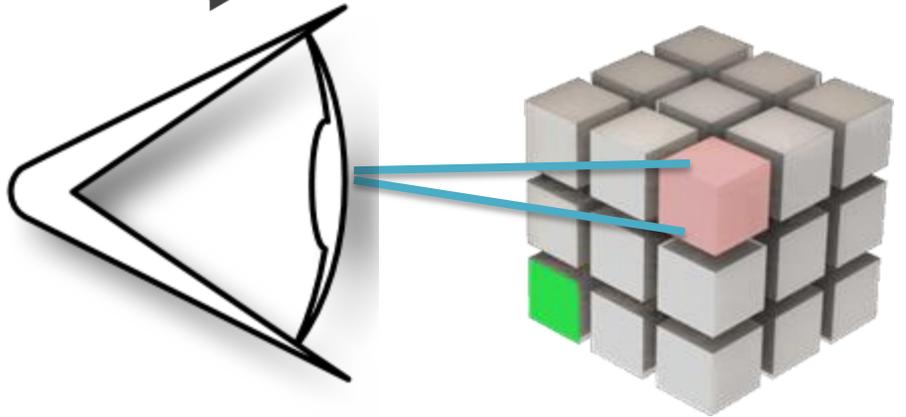
## Resource Utilization



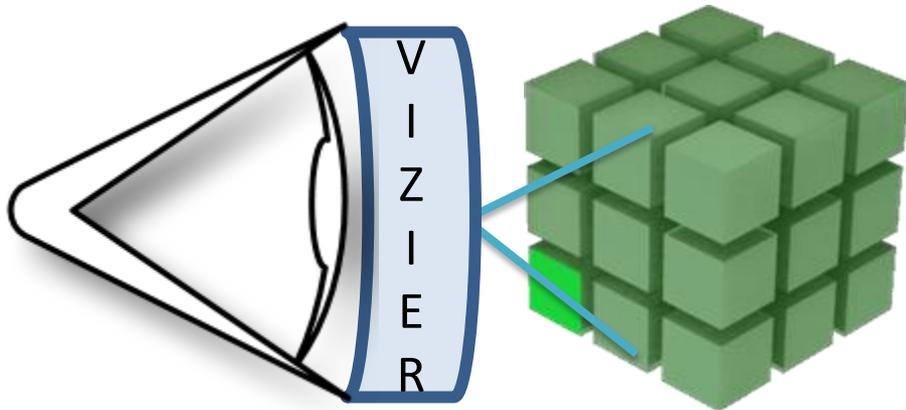
- Process RAM Requirement
  - 60 MB Memory SQL
  - 90 MB Memory Oracle
- 0 MB Disk Utilization
- 0 MB Database Utilization



## The Vizier™ Difference



Because your application 'sees' only a tiny subset of data at a time, it may miss the data that you are actually looking for (represented by the bright green cube)

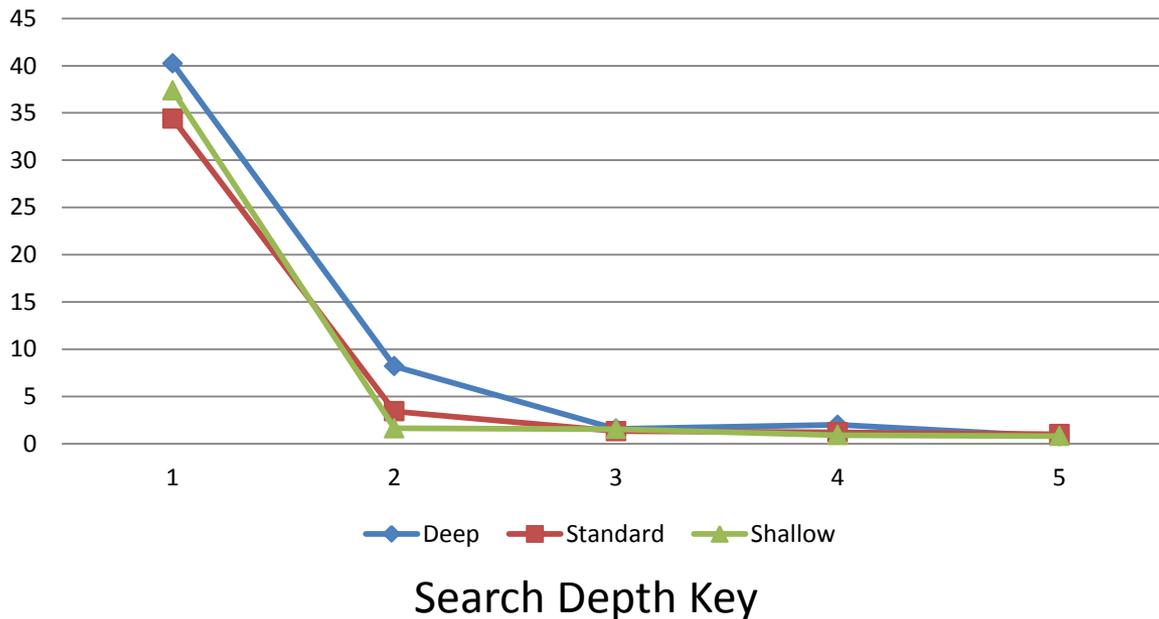


With Vizier™, your application can see much more, which increases your chances of returning what you want



## Relationship of Correct Entry to Response Time

**Response Time for Correct Inputs (in seconds)**



Correct argument entries by the user shrink the search space, which reduces search times

The size of the search space is the determinant of response time, not the size of the database itself



## Demonstration Hardware Specs

- Database Server:
  - 10Gb RAM
  - AMD Athlon 7750 Dual Core processor 2.70GHz
  - 1 TB disk space
- Web Application Server
  - 4Gb RAM
  - Intel Core 2 Quad CPU Q9550 2.83GHz and 2.83GHz
  - 500 Gb disk space