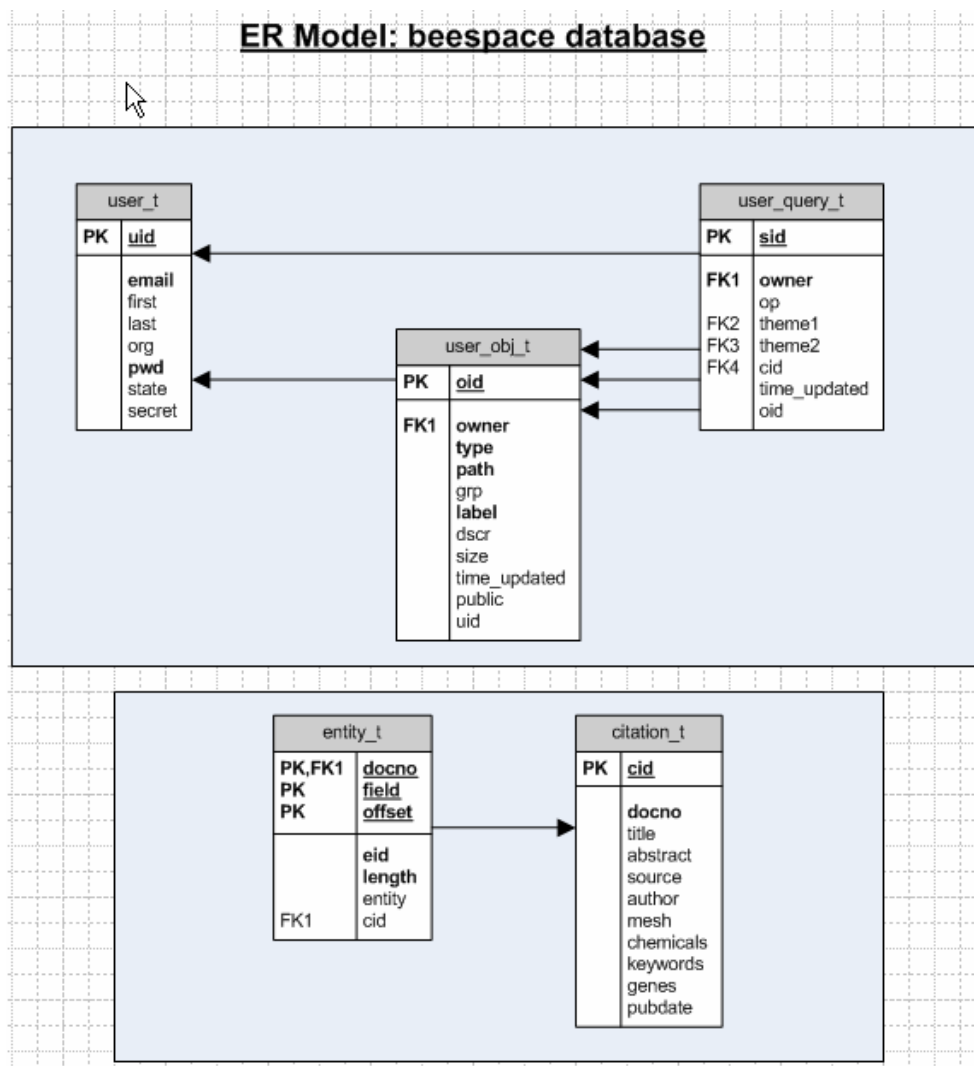


# Entity-Relationship Model of the BeeSpace Navigator Database – September 2007



This is the database scheme for BeeSpace v3. The 3 relational tables at the top are the session dynamic values; they are filled in as the user navigates through the system. The tables left to right store values for the user logins, the user objects (spaces and regions), and the user queries. The 2 relational tables at the bottom are the collection static values, they are filled in as new documents are imported into the system. The left table is the parsed entities from the contents while the right table is the documents with metadata. Adding entities is easy, requiring only an alter table and an add column command.

The Java web application with the user interface also retrieves from other sources besides the MySQL database. The full-text indexing for search is supported by the Lemur/Indri toolkit. Temporary indexes in the local file system, such as the semantic graph for mutual information, are only dynamically generated, not permanently stored.

## User Table

The User Table contains the directory of BeeSpace users. It stores both visible and hidden fields of a user's account.

The screenshot shows a database management interface for the 'user\_t' table. On the left, a table diagram shows columns: uid (PK), email, first, last, org, pwd, state, and secret. The 'Categories' pane on the right is set to 'Definition'. Below this, a table lists the column definitions:

Physical Name	Data Type	Req'd	PK	
uid	INTEGER	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	uid is auto-generated user id (PK)
email	VARCHAR(100)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	unique email of user
first	VARCHAR(40)	<input type="checkbox"/>	<input type="checkbox"/>	given name
last	VARCHAR(40)	<input type="checkbox"/>	<input type="checkbox"/>	family name
org	VARCHAR(100)	<input type="checkbox"/>	<input type="checkbox"/>	user's organization
pwd	VARCHAR(20)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	password
state	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	user's state variable
secret	VARCHAR(40)	<input type="checkbox"/>	<input type="checkbox"/>	user's secret

Column definitions:  
uid is auto-generated user id (PK)  
unique email of user  
given name  
family name  
user's organization  
password  
user's state variable  
user's secret

The screenshot shows the 'Notes' property for the 'user\_t' table. The 'Categories' pane on the left is set to 'Notes'. The 'Notes' field contains the following text:

Directory of beespace users; stores both visible and hidden fields of a user's account.

## User Object Table

The User Object Table stores user objects (e.g. regions/spaces) and their metadata.

Physical name: user\_obj\_t  
 Conceptual name: User Object Table  
 Name space: beespace  
 Owner:  
 Source database:

Physical Name	Data Type	Req'd	PK	
oid	INTEGER	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	oid is auto-generated object id (PK)
owner	INTEGER	<input checked="" type="checkbox"/>	<input type="checkbox"/>	owner of this operation
type	CHAR(1)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	type code of object
path	VARCHAR(255)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	link to object
grp	VARCHAR(100)	<input type="checkbox"/>	<input type="checkbox"/>	user-assigned group
label	VARCHAR(100)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	user-assigned label
dscr	VARCHAR(255)	<input type="checkbox"/>	<input type="checkbox"/>	user-assigned description text
size	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	object size
time_updated	LONG	<input type="checkbox"/>	<input type="checkbox"/>	last create/update timestamp
public	BIT	<input type="checkbox"/>	<input type="checkbox"/>	is-public flag
uid	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	uid is of User Object Table

Show:  Portable data type  Physical data type (Microsoft Access)

Column definitions:

oid is auto-generated object id (PK)

owner of this operation

type code of object

link to object

user-assigned group

user-assigned label

user-assigned description text

object size

last create/update timestamp

is-public flag

uid is of User Object Table

Categories:

Notes:  
 Stores user objects (e.g. regions/spaces) and their metadata.

## User Query Table

The User Query Table stores user queries, which could be used for displaying or re-executing history of user operations, etc.

The screenshot shows a database management interface for the 'user\_query\_t' table. On the left, a table structure is displayed with columns: sid (PK), owner (FK1), op (FK2), theme1 (FK3), theme2 (FK4), cid, time\_updated, and oid. The main area shows the 'Columns' category selected, displaying a table of column definitions.

Physical Name	Data Type	Req'd	PK	Description
sid	INTEGER	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	sid is auto-generated query id (PK)
owner	INTEGER	<input checked="" type="checkbox"/>	<input type="checkbox"/>	owner of this operation
op	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	op code
theme1	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	first operand
theme2	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	second operand
cid	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	collection id
time_updated	LONG	<input type="checkbox"/>	<input type="checkbox"/>	last create/update timestamp
oid	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	oid is of User Query Table

At the bottom, there are radio buttons for 'Portable data type' and 'Physical data type' (selected), with '(Microsoft Access)' noted next to it.

Column descriptions:

sid is auto-generated query id (PK)

owner of this operation

op code

first operand

second operand

collection id

last create/update timestamp

oid is of User Query Table

The screenshot shows the 'Notes' section of the database management tool. The 'Notes' field contains the text: 'Stores user queries, which could be used for displaying or re-executing history of user operations, etc.'

## Entity Table

The Entity Table stores entities found from Jing's NER program. Fields offset and length are character based metrics for identifying the entity text. The entity field is rather redundant and was used for backtracing/debugging primarily.

entity\_t

PK,FK1	docno
PK	field
PK	offset
	eid
	length
	entity
FK1	cid

Categories:

- Definition
- Columns
- Primary ID
- Indexes
- Triggers
- Check
- Extended
- Notes

Physical name: entity\_t

Conceptual name: NER Table

Name space: beespace

Owner:

Source database:

Categories:

- Definition
- Columns
- Primary ID
- Indexes
- Triggers
- Check
- Extended
- Notes

Physical Name	Data Type	Req'd	PK	Notes
eid	INTEGER	<input checked="" type="checkbox"/>	<input type="checkbox"/>	eid is auto-generated entity record id (PK)
docno	VARCHAR(100)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	unique email of user
field	VARCHAR(40)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	field in record entity is found
offset	INTEGER	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	character offset into field of start pos
length	INTEGER	<input checked="" type="checkbox"/>	<input type="checkbox"/>	character length of entity text
entity	VARCHAR(255)	<input type="checkbox"/>	<input type="checkbox"/>	the actual entity text (mainly for backtracing)
cid	INTEGER	<input type="checkbox"/>	<input type="checkbox"/>	cid is of NER Table

Column definitions:

eid is auto-generated entity record id (PK)

unique email of user

field in record entity is found

character offset into field of start pos

character length of entity text

the actual entity text (mainly for backtracing)

cid is of NER Table

Categories:

- Definition
- Columns
- Primary ID
- Indexes
- Triggers
- Check
- Extended
- Notes

Notes:

Stores entities found from Jing's NER program. Fields offset and length are character based metrics for identifying the entity text. The entity field is rather redundant and was used for backtracing/debugging primarily.

# Citation Table

The Citation Table stores the normalized contents of a beespace citation record. Note: the genes field corresponds to human-annotations that come directly from PubMed for Biosis and hence are rather sparse.

Categories:

- Definition
- Columns
- Primary ID
- Indexes
- Triggers
- Check
- Extended
- Notes

Physical name: citation\_t  
 Conceptual name: Citation Table  
 Name space: beespace  
 Owner:  
 Source database:

Physical Name	Data Type	Req'd	PK	Notes
cid	INTEGER	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	cid is auto-generated citation db id (PK)
docno	VARCHAR(100)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	citation external docno
title	TEXT(1000)	<input type="checkbox"/>	<input type="checkbox"/>	title of citation
abstract	TEXT(1000)	<input type="checkbox"/>	<input type="checkbox"/>	abstract of citation
source	TEXT(1000)	<input type="checkbox"/>	<input type="checkbox"/>	publication source of citation
author	TEXT(1000)	<input type="checkbox"/>	<input type="checkbox"/>	author field of citation
mesh	TEXT(1000)	<input type="checkbox"/>	<input type="checkbox"/>	mesh field of citation
chemicals	TEXT(1000)	<input type="checkbox"/>	<input type="checkbox"/>	chemicals field of citation
keywords	TEXT(1000)	<input type="checkbox"/>	<input type="checkbox"/>	keywords field of citation
genes	TEXT(1000)	<input type="checkbox"/>	<input type="checkbox"/>	genes field of citation
pubdate	VARCHAR(100)	<input type="checkbox"/>	<input type="checkbox"/>	publicatoin date

Column definitions:

- cid is auto-generated citation db id (PK)
- citation external docno
- title of citation
- abstract of citation
- publication source of citation
- author field of citation
- mesh field of citation
- chemicals field of citation
- keywords field of citation
- genes field of citation
- publication date

Categories:

- Definition
- Columns
- Primary ID
- Indexes
- Triggers
- Check
- Extended
- Notes

Notes:  
 Stores the normalized contents of a beespace citation record. Note: the genes field corresponds to human-annotations that come directly from PubMed for Biosis and hence are rather sparse.