HTSQL - A Navigational Query Language For Relational Databases

Charles Tirrell, Clark Evans, Kyrylo Simonov

PROMETHEUS | RESEARCH™

Abstract

HTSQL is a high-level navigational query language and web-service gateway used by medical researchers to answer complex data inquiries in a clinical setting. HTSQL has been in development for 7 years and in the forthcoming 2.4 release adds a really superb ETL translation mechanism to handle integration of related data-sets and the construction of derived databases. This poster will introduce HTSQL and review its broad applicability to scientific computing.

To Learn More...

htsql.org

rexdb.org

A Query Language

HTSQL is a complete query language featuring automated linking, aggregation, projections, filters, macros, a compositional syntax, and a full set of data types and functions. It is designed for data analysts and other accidental programmers who have complex business inquiries to solve and need a productive tool to write and share database queries.

HTSQL Queries are Web URLS

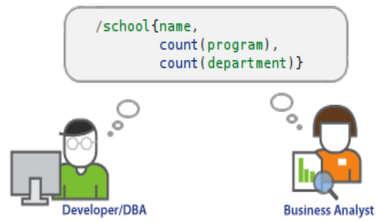
```
http://www.example.org
/protocol
.filter(
        is_true(study.is_open)&
        exists(participation?individual.sex='male')
.select(
        code,
        title,
        count(requirement))
/:csv
```

Supports Common Output Formats

HTML, XML , JSON, YAML, CSV, R

Helps Technical/Non-technical Folk Communicate

HTSQL is used for collaboration among researchers, data managers, and application developers. HTSQL queries can be emailed, embedded in reports, and included in feature requests.



An ETL Transform Language

Common tasks in scientific research requires leveraging data from multiple public and private sources. However, data from different sources may be difficult to link because data are organized differently, of inconsistent/poor quality or lack appropriate identifiers. The standard solution to these challenges is to use an Extract, Transform and Load process (ETL), which pulls the data out of the original data source, changes its structure and content, then stores it in the target database. However, typical ETL processes tend to be brittle, expensive, and difficult to run.

Example Transform

```
/source_patient_table
.define(
    $code := memberid :string,
    $mbrsex :=
        if(
        mbrsex='Female','female',
        mbrsex='Male','male',
        mbrsex='NotKnown','not-known')
)
{
    code := $code,
    sex := $mbrsex
}
:as target_patient_table
/:gateway_to_sqlitedb
/:merge
```

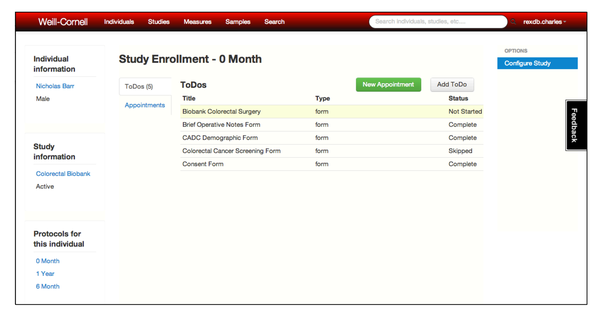
The HTSQL-based ETL mechanism allows researchers to mount CSVs, the common data format in human behavioral and biomedical research, into a temporary SQLite database or to import an existing database and add HTSQL as a data access layer. They can then write HTSQL queries that define result sets in the same format as the target table for easy merging of the data.

A Data Access Layer for Scientific Application Development

The Research Exchange Database (RexDB®) is an extensible, web-native software platform that helps researchers securely collect, manage and share human behavioral and biomedical data. Unlike most data centralization tools, RexDB embraces the flexibility required to support the needs of dynamic scientific collaborations. RexDB is built on top of sustainable open-source components (Python, PostgreSQL and HTSQL), ensuring that research data will never be marooned in a proprietary format or a legacy system.

HTSQL is a web service that accepts queries as URLs, returning results formatted as HTML, JSON, CSV or XML. With HTSQL, databases can be accessed, secured, cached, and integrated using standard web technologies. Combining HTSQL with HTRAF, a Javascript framework for embedding data into HTML pages, enabled the rapid and flexible development of the RexDB Application.

RexDB Study Enrollment Page



Example Data Search in RexDB with HTSQL

"Across all studies in our center, show me the BAPQ, SRS, ADOS and ADI-R scores for everyone who has a verbal IQ above 75, no known genetic lesions at 16p11.2, and whose mother had a history of autoimmune disorders. OK, now limit the data set to those participants for whom we already have a live-cell-line biospecimen or those I have consent to re-contact."

```
/individual
.select(identity.*,bapq.*,srs.*,ados.*,adir.*)
.filter(
        verbaliq>75&
        !exists(lesion?loc='16p11.2)&
        mother.exists(
            condition?disorder='autoimmune-disorder')&
        exists(sample?sample_type='cell-line')&
        exists(consent?consent_type='recontact'))
```

Example HTSQL/HTRAF Div

```
<div id="object" data-ref="search id  " data-htsql="/study {_id:hidden,
title, count(protocol) :as Protocols,
count(protocol.protocol_participation) :as Participants } .filter(_id~
$search|is_null($search))" data-widget="grid" class="tablemain"></div>
```