

Freebase: A Collaboratively Created Graph Database For Structuring Human Knowledge

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, Jamie Taylor
Metaweb Technologies, Inc.
631 Howard St., San Francisco, CA 94105

ABSTRACT

Freebase is a practical, scalable tuple database used to structure general human knowledge. The data in Freebase is collaboratively created, structured, and maintained. Freebase currently contains more than 125,000,000 tuples, more than 4000 types, and more than 7000 properties. Public read/write access to Freebase is allowed through an HTTP-based graph-query API using the Metaweb Query Language (MQL) as a data query and manipulation language. MQL provides an easy-to-use object-oriented interface to the tuple data in Freebase and is designed to facilitate the creation of collaborative, Web-based data-oriented applications.

Categories and Subject Descriptors

H.3.4 Systems and Software, Information Networks; E.1 Data Structures, Graphs and networks.

General Terms

Design, Human Factors, Languages

1. INTRODUCTION

Freebase[3] is a database system designed to be a public repository of the world's knowledge. Its design is inspired by broadly used information communities such as The Semantic Web[2] and Wikipedia[1]. As such, it must support highly diverse and heterogeneous data simultaneously with high scalability.

Traditionally structured databases have been centrally controlled, with schemas created and modified only by a set of trusted administrators. Support for structural diversity is often difficult in such architectures. At the other end of the spectrum, popular wikis (e.g. Wikipedia) are semi-structured document repositories of great semantic heterogeneity, but with few tools providing structured query capabilities.

Freebase tries to merge the scalability of structured databases with the diversity of collaborative wikis into a practical, scalable, database of structured general human knowledge. The

design emphasizes practicality, and a substantial effort has been made to make Freebase immediately useful, especially to laymen beyond the research and technical communities.

The key components of Freebase are:

- **A scalable Tuple Store** that has built-in attribution and support for total “undo” through an integrated versioning mechanism. Together, these features facilitate large-scale collaborative data creation and maintenance.
- **An HTTP/JSON-Based API** where read and write queries and their results are formulated using the “Metaweb Query Language”[4] (MQL). MQL is an easy to use, object-oriented query language with a tree based result structure of objects. It includes dynamic schema support without the need for a DDL, path based node naming and idempotent transaction-less write support. Important MQL features include mixing structural data matching with approximate string matching of literals, cursors, and flat semantics of all data, which makes mixing of data and metadata easy.
- **A Lightweight, Collaborative Typing System.** Pervasive in the UI, API, and data itself is the Freebase typing system. It is a loose collection of structuring mechanisms and conventions, rather than a rigid systems of ontologies and representations. The type system has no intrinsic canonical view of data. Conflicting and contradictory types and properties may exist simultaneously in order to reflect users' differing opinions and understanding.
- **A Large, Diverse Data Set.** Freebase already has more than 100 million asserts about topics spread over 4000 types, including people, media, locations, and many others. This data is available through a Creative Commons license for almost any use.
- **A Philosophy of “Complete Normalization”.** Unlike in some databases, entities in Freebase are intended to be explicitly canonicalized. That is, there should be only one GUID in Freebase representing each real world entity, topic, or concept.

2. DEMONSTRATION OF FREEBASE

We will demonstrate how Freebase and MQL enable the accumulation of structured information in a manner that increases quality as well as quantity of data. The MQL

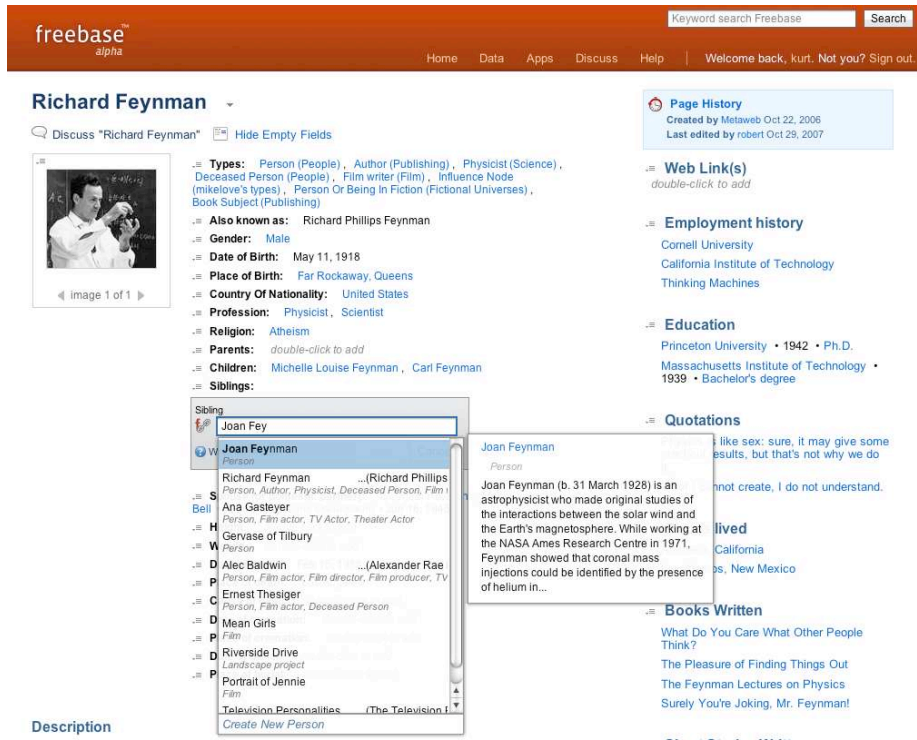


Figure 1: This is the main Topic Display page in Freebase. The user is editing the “Richard Feynman” topic. Autosuggestion helps the user add the sibling property by linking to an existing object, rather than create an unnecessary duplicate.

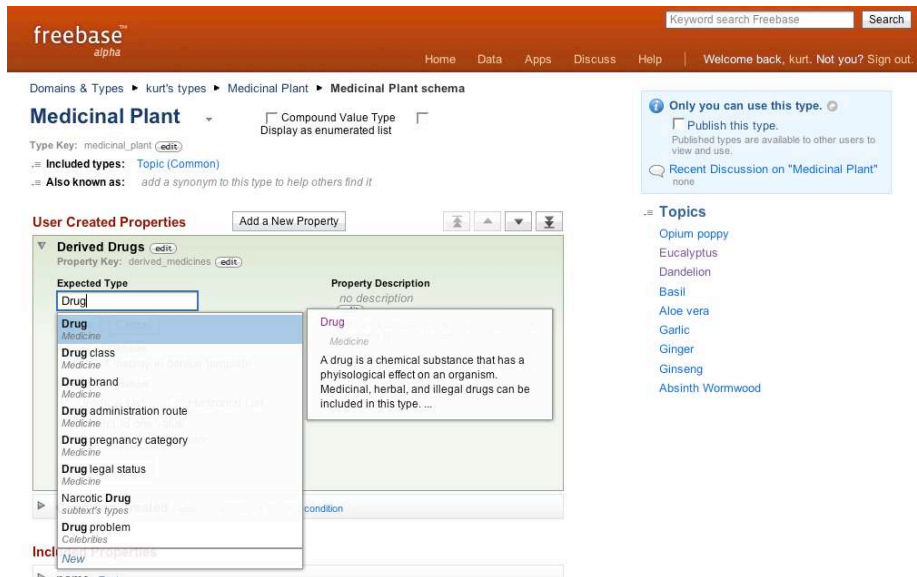


Figure 2: This is the Schema Editor used for schema creation and evolution. The user is editing the “Medicinal Plant” schema to add a “Derived Drug” property that expects an instance of “Drug” to be at the other end of the link when instances of this property are created.

freebase™
S A N D B O X
alpha

Keyword search Freebase Search

Home Data Apps Discuss Help Welcome back, kurt. Not you? Sign out.

Domains & Types ▶ Film ▶ Film actor

Import a List of Film actors

Step 2 of 3: Reconcile Your List of Film actors with Existing Film actors on Freebase

Some items in your list may already exist on Freebase, but aren't yet linked to Film actor. Reconcile this by comparing items in your list with existing items and deciding which to merge, skip, or add.


5 items were automerged.
1 of 6 left to manually reconcile.

Start Over Modify List Import List

List Item	Status
Bruce Campbell	likely merge
Jason Biggs	merge
Lucas Cain	pending
Tyler Hilton	likely merge
Don Johnson	likely merge
Val Kilmer	merge

Bruce Campbell will be merged with:

Confirm Merge **Bruce Campbell**

 *Topic, Person, Film actor, Film producer, Film director, Film writer, TV Actor, Author*
For the former baseball player of the same name, see Bruce Campbell (baseball). For the Home and Away character of the same name, see Bruce Campbell (Home and Away) Bruce Lorne Campbell (born June 22 1958) is an American actor, producer, writer and director. He is best known for his starring role...

Do something different with **Bruce Campbell**:

Merge With **Bruce Campbell**

Topic, Person
Bruce Campbell (October 20, 1909 - June 17, 1995) was a professional baseball player from 1930 to 1942. Campbell began his career with the Chicago White Sox, but had very little playing time in the major leagues. In 1932, Campbell was traded from the White Sox to the St. Louis Browns with Bump...

Merge With **Keith Bruce Campbell**

Figure 3: The Mass Typer application lets the user submit a list of topic names that are semi-automatically reconciled during interaction with the user. The figure here shows this application as applied to an imported list of film actors.

language will be described through examples and demonstration on the live database. Several existing, practical walkthroughs and Web-based Freebase applications will be shown with the MQL they use, including:

- Data entry, shown by *Topic Display* (Figure #1)
- Schema creation and evolution, shown by *Schema Editor*: (Figure #2).
- Rapid Data Traversal
- Data evaluation and reconciliation, shown by *Mass Typer*: (Figure #3)
- Collaborative data analysis and repair
- MQL query creation and refinement

Freebase is intended to be a perpetually available public service on the Web. Readers interested in better understanding Freebase and MQL are encouraged to explore on their own at www.freebase.com.

3. REFERENCES

- [1] "wikipedia (n.d.). wikipedia: The free encyclopedia. on the internet. <http://www.wikipedia.org>.
- [2] T. Berners-Lee, J. Hendler, and O. Lassila. "the semantic web". *Scientific American*, 284:34-43, 5 2001.
- [3] K. Bollacker, P. Tufts, T. Pierce, and R. Cook. A Platform for Scalable, Collaborative, Structured Information Integration. In *Sixth International Workshop on Information Integration on the Web*. Association for the Advancement of Artificial Intelligence, July 2007.
- [4] D. Flanagan. *Developing Metaweb-enabled Web Applications*. Metaweb Technologies, 2007.