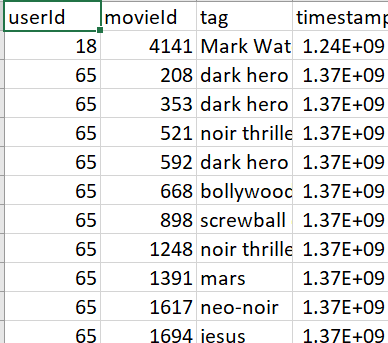
https://grouplens.org/datasets/movielens/

**Tags Data File Structure (tags.csv)**

-----------------------------------

All tags are contained in the file `tags.csv`. Each line of this file after the header row represents one tag applied to one movie by one user, and has the following format:

userId,movieId,tag,timestamp



The lines within this file are ordered first by userId, then, within user, by movieId.

Tags are user-generated metadata about movies. Each tag is typically a single word or short phrase. The meaning, value, and purpose of a particular tag is determined by each user.

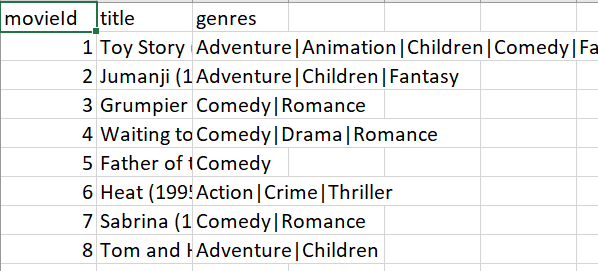
Timestamps represent seconds since midnight Coordinated Universal Time (UTC) of January 1, 1970.

**Movies Data File Structure (movies.csv)**

---------------------------------------

Movie information is contained in the file `movies.csv`. Each line of this file after the header row represents one movie, and has the following format:

movieId,title,genres



Movie titles are entered manually or imported from <https://www.themoviedb.org/>, and include the year of release in parentheses. Errors and inconsistencies may exist in these titles.

Genres are a pipe-separated list, and are selected from the following:

\* Action

\* Adventure

\* Animation

\* Children's

\* Comedy

\* Crime

\* Documentary

\* Drama

\* Fantasy

\* Film-Noir

\* Horror

\* Musical

\* Mystery

\* Romance

\* Sci-Fi

\* Thriller

\* War

\* Western

\* (no genres listed)

**Links Data File Structure (links.csv)**

---------------------------------------

Identifiers that can be used to link to other sources of movie data are contained in the file `links.csv`. Each line of this file after the header row represents one movie, and has the following format:

movieId,imdbId,tmdbId

movieId is an identifier for movies used by <https://movielens.org>. E.g., the movie Toy Story has the link <https://movielens.org/movies/1>.

imdbId is an identifier for movies used by <http://www.imdb.com>. E.g., the movie Toy Story has the link <http://www.imdb.com/title/tt0114709/>.

tmdbId is an identifier for movies used by <https://www.themoviedb.org>. E.g., the movie Toy Story has the link <https://www.themoviedb.org/movie/862>.

Use of the resources listed above is subject to the terms of each provider.

**Tag Genome (genome-scores.csv and genome-tags.csv)**

-------------------------------------------------

This data set includes a current copy of the Tag Genome.

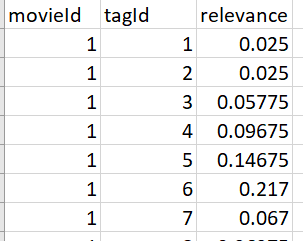
[genome-paper]: <http://files.grouplens.org/papers/tag_genome.pdf>

The tag genome is a data structure that contains tag relevance scores for movies. The structure is a dense matrix: each movie in the genome has a value for \*every\* tag in the genome.

As described in [this article][genome-paper], the tag genome encodes how strongly movies exhibit particular properties represented by tags (atmospheric, thought-provoking, realistic, etc.). The tag genome was computed using a machine learning algorithm on user-contributed content including tags, ratings, and textual reviews.

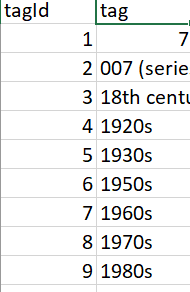
The genome is split into two files. The file **`genome-scores.csv**` contains movie-tag relevance data in the following format:

movieId,tagId,relevance



The second file, `**genome-tags.csv`,** provides the tag descriptions for the tag IDs in the genome file, in the following format:

tagId,tag



The `tagId` values are generated when the data set is exported, so they may vary from version to version of the MovieLens data sets.

Cross-Validation

----------------

Prior versions of the MovieLens dataset included either pre-computed cross-folds or scripts to perform this computation. We no longer bundle either of these features with the dataset, since most modern toolkits provide this as a built-in feature. If you wish to learn about standard approaches to cross-fold computation in the context of recommender systems evaluation, see [LensKit](http://lenskit.org) for tools, documentation, and open-source code examples.