

Curated Open Repository of Progressive Song Entries (CORPSE)

This document covers many components of the CORPSE project including:

- 1. Information about the project, what is included and how to compile it
- 2. How to use the CORPSE Explorer
- 3. Which bands and albums are included (Roughly 415 Albums)
- 4. How the entries are tagged

CORPSE Author

Ty Gordon

Most Recent Update

12/10/2024

Number of Artists Included:

101

Number of Albums Included:

422

Number of Tracks Included:

3,741

Number of Words Included:

11,480

Table of Contents

1. [Overview](#)
2. [Usage Guide](#)
3. [Artist List](#)
 - [The Big Seven](#)
 - [Secondary Bands](#)
 - [Solo Artists](#)
 - [Canturbury Scene](#)
 - [Fusion](#)
 - [Krautrock](#)
 - [RPI](#)
 - [International](#)
 - [Artists to Include](#)
4. [Tagging](#)

OVERVIEW

What Is This?

This is the Curated Open Repository of Progressive Song Entries (CORPSE), which is a project developed for the University of Kentucky Corpus Linguistics course LIN-510. At its core, CORPSE is a collection of XML formatted song titles sourced from progarchives.com within the years 1966 to 1984. It comes with an explorer utility, which amounts to a Graphical User Interface (GUI) designed to allow any user the ability to search the database in a simple manner (unlike CQP, RegEx, etc).

Progressive Rock is a term used to describe a genre of music that was prevalent in the late 60s, to the early 80s. It is usually described as a more complicated and eclectic form of popular music with many influences from genres such as Jazz, Classical, Blues, Folk Baroque, Soul, R&B, Electronic, Experimental, and Rock. This corpus only includes the "classic" era of Progressive Rock, and deliberately excludes newer forms passed 1984 such as Progressive Metal, Neo-Prog, etc. Progressive Rock itself can be described with various sub-genres. The following links may be helpful for understanding the genre.

[Prog Archives Home](#)

[What is Progressive Rock?](#)

[Top 100 Prog Albums 1966-1984](#)

Purpose

The corpus was built to answer questions about the stylistic conventions of progressive rock titles. The contents of the CORPSE corpus are far too limited to make any generalizations about english or how english is generally used. It is, however, adequate for answering questions about progressive rock song titles since it encompasses a significant portion of all progressive rock titles from the classic era. The querying systems makes use of abundant meta-data to compensate for the lack of large sums of data, and accurate tagging. Questions such as "Do prog songs generally get shorter over time" or "Do long prog songs have shorter titles" are the kinds of questions that this corpus aspires to help answer.

What's Included

The CORPSE Corpus consists of two main tools and the data that makes up the corpus (both raw html and formatted xml files). The tools CORPSE Maker and CORPSE Explorer are made up of the following subcomponents:

CORPSE MAKER

1. README.md (This file) - To be used as the source of the URLs.
2. curl_maker.py - A python program that builds a series of curl commands to fetch web data.
3. corpus_curl.ps1 - The powershell script output from curl_maker.py. When run it executes the curl commands.
4. track_listing_scraper.py - A python program that takes the HTML pages from curl and converts it to formatted XML files to be used in the corpus.
5. /Albums - A folder of all of the HTML pages used for the corpus.
6. /Data - A folder of each album represented in an XML format.
7. /Images - A folder with some images used for the GUI and its development.
8. full_corpus.xml - An XML file with all of the data. This is used by the CORPSE Explorer.

CORPSE EXPLORER

1. corpus_explorer.py - A PyQt application that performs searches over the XML data.
2. corpus_gui.ui - A QT ui file used for render the explorer GUI.
3. resources.qrc - a qrc file used by QT to load images, etc.

OTHER

1. 250_finder.py - A python program that finds all of the artists from the list of the top 250 albums from prog archives.

2. top-250.html and top-250.txt - The input and output of 250_finder.py.
3. manual_scraper.py - A script to convert manually typed albums into lemmas, POS, etc
4. manual_albums.txt - The output of the manual scraper
5. problematic_albums.txt - A list of problematic albums

What's Needed

These resources are necessary for re-assembling the corpus:

- Microsoft Windows 10/11 OS (linux/mac are untested)
 - Curl for Windows (for curl_maker.py to work)
 - Python, pip (for .py programs)
 - NLTK (use pip) (used for POS tagging)
 - deep-translator (use pip) (used for translation)
 - QT and pyqt (use pip) (used for GUI)
 - Pyinstaller for making .exe
 - ==> Use this command:
- ```
pyinstaller.exe --onefile --name="CORPSE Explorer" --add-data="full_corpus.xml:." --add-data="resources.qrc:." --add-data="corpus_gui.ui:." --add-data="Images/Images" --icon="Images/uk_icon.ico" corpus_explorer.py
```

Otherwise, just running the CORPSE\_Explorer.exe should work with no added software.

(full\_corpus.xml, resource.qrc, and corpus\_gui.ui need to be in the same directory to work)

## Limitations

Known Bugs:

- Track names with a period followed by a space do not play nicely with the POS tagger: the tagger tends to eat them.
- XML does not like the '&' character, so all instances were replaced with "and"
- CURL does not like some characters such as '/', which had to be replaced with ' ',
- Many special characters (such as those of the cyrillic variety, or those with uncommon accent markers) are not compatible with the character encoding used. They are either left unchanged, or hand-corrected.
- The format of the sourced HTML pages can change, and did in-fact change upon the second corpus compilation on Nov 15th. The spacing in the source code for HTML elements may change over time as was the case here. New features could also have been added to the website, which could change what lines are needed for data extraction.
- Albums with multiple sub-tracks of the same index may show the wrong track since they share indexes.

- Upon compilation, the corpus may have some ratings and/or indexes set to a string (since the data is inconsistent). This results in a crash when using the explorer. These have to be manually corrected for now.
- Some tracks may have an asterisk to indicate additional information. Sometimes these tracks get filtered out such as "Incantations Part 3\*" from Mike Oldfield on the album "Incantations".
- QT spews font errors for some reason. This does not impact the functionality of the program

#### Design Flaws:

- Many albums are excluded for the following reasons:
  - > They do not fit within the time frame (1966-1984).  
Ex. "Invisible Touch" by Genesis
  - > They are under a different name.  
Ex. Magma's "Tristan et Iseult" ("Wurdah Itah")
  - > The album has a very poor relative rating AND is at the end of the group's "golden era".  
(This is done because many groups tended to switch genres away from Prog rock, so these albums are typically not representative of the genre this corpus encompasses)  
Ex. Jethro Tull's "Under Wraps"
  - > The respective group is niche or borderline prog and only gets one album, with the rest omitted.  
Ex. Return to Forever's albums that are not "Romantic Warrior"
  - > The group was not discovered/known about at corpus compiling time.
- Tracks from foreign groups may need to be translated by the auto-translator. This is prone to mistakes and may influence the style of foreign titles (ie. romance languages adding a surplus of the word "of", etc).
- Albums with more than 12 sub-tracks under one main track are not supported. a-1 and i-xii are the only sub-track alphabet index characters that are converted into a numeric index.
- The automatic XML generation makes any sub-track listed with an index of "-" have no proper index.  
THESE NEED TO BE FIXED MANUALLY.
- Some album entries do not follow the common format and are problematic. Some examples are:
  - Joe's Garage Act, I - Frank Zappa  
=> Has no track times
  - Roxy & Elsewhere - Frank Zappa  
=> Misspelt "Total Time"
  - Who's Next - The Who  
=> No "Total Time"
  - Private Parts and Pieces III - Anthony Philips  
=> No period after numerals
- Cyrillic characters do not work with the translation step currently, so entries from Aftograf are excluded
- The "chain-search" feature

### Flaws With 3rd Party Utilities:

- The lemmatizer used was either incorrectly used or is generally bad. Many lemmas are wrong such as:

was -> wa  
does -> doe  
has -> ha

These are not corrected.

- The POS tagger is also often wrong. Most song titles are capitalized, and are thus tagged as NNS.  
- The translator translates everything. This wasn't the original plan, but this is how it is.

THE FOLLOWING ALBUMS ARE PROBLEMATIC AND NEED TO BE HAND-CORRECTED UPON CORPUS COMPILATION:

(List Format: [Album Title] - [Artist], '\n' '\t' '\t' => [Notes on why the title is an issue])

- Joe's Garage Act, I - Frank Zappa ✓  
=> Tracks lack numbers
- Roxy & Elsewhere - Frank Zappa ✓  
=> Misspelt "Total Time"
- Who's Next - The Who ✓  
=> No "Total Time"  
=> Track Times use a period, not a colon
- Private Parts and Pieces III - Anthony Philips ✓  
=> No period after numerals
- More - Pink Floyd ✓  
=> Unknown failure
- Per Un Amico - Premiata Forneria Marconi ✓  
=> Unknown failure
- Magick Brother - Gong ✓  
=> Side names may be the issue
- Klaus Schultz - Moondawn ✓  
=> Unkown failure
- Jatekok - East ✓  
=> Non-Latin characters, translation provided
- Huseg - East ✓  
=> Non-Latin characters, translation provided
- Avtograf - 1 ✓  
=> Cyrillic characters, translation provided  
=> MANUALLY CHANGE DATE TO 1982 FOR ACCURACY
- Avtograf - 2 ✓  
=> No Prog archives entry. Use the youtube link  
=> This album was released informally as a "magnito-album" <<Магнитоальбом>> with an imprecise track-listing. The Youtube version will be used as the source for the data.
- Magma  
=> This group invented a con-lang in order to foil my corpus plans. The special characters do not play nice with the scraping tools and thus may appear as unknown characters.

## Lessons Learned

A lot had to be learned to make this project what it is since many unfamiliar tools were needed. Here is a list of the skills gained and practiced during development:

- Python Programming
  - => Reading Files
  - => Writing to Files
  - => Using 3rd party libraries such as:
    - NLTK
    - Deep Translator
    - Element Tree
    - lxml
    - Pyqt 6
- QT 6
  - => Designing a functional GUI
  - => Implementing a GUI in QT
  - => Connecting GUI elements to python functions
  - => Using QT signals
- Web Scraping
  - => Navigating HTML code
  - => Using Wget and Curl to fetch webpages
  - => Using (lots of) RegEx to clean the html data
- XML Formatting
  - => Designing an XML database
  - => Formatting input into XML tags
- XML Querying
  - => Building Xpath expressions
  - => Using Xpath to find entries
  - => Filtering by XML tags
- Programming Practices
  - => Implementing parallel processing
  - => Building and querying databases
  - => Documenting a project
  - => Using markdown
  - => Using version control (Git)
  - => Using Gen AI resources to help implement a solution
  - => Coming up with great names and acronyms
- Corpus Creation
  - => Designing a corpus
  - => Metadata collection
  - => Translation
  - => Lemmatization
  - => PoS tagging
- Corpus Investigation
  - => Making informative queries
  - => Using statistics to analyze patterns
  - => Writing investigation reports

## Future Work

This project is by no means complete. Several features and ideas are left incomplete, non-functional or untouched. Even the functionality provided is inherently flawed, and could be reworked for a better product.

Here are some features that were planned to be added if development were to continue:

- Chain-Search
  - ==> Almost implemented, but scrapped due to time and implementation difficulty. Would allow the user to "chain" or "stack" queries whereby the results of one query would be stored as a "subset-corpus" and be fed as the input for a subsequent query.
- Multi-Word Search
  - ==> Instead of, or alongside the "chained-search" feature, support for searching multiple consecutive words would be a great feature to have.
- Track-Type Tag
  - ==> Some implementation (GUI). Would allow the user to sort by tracks that are sub-tracks or parts of a larger track. Or lets them filter them out if undesired. These tags are not present in full\_corpus.xml
- Search by Last Word
  - ==> Implement a way to search by the last index.
- "Misc" Tags
  - ==> Tag the titles for extra meta information such as "has-pun", "comedic", "explicit language", "contains reference", "non-semantic phrase", etc.
- Semantic Tagging
  - ==> Add a semantic tagger to add these tags within the track scraper.
- Semantic Trees
  - ==> Add a system to store the titles as trees and visualize them in other windows.
- N-Gram Search
  - ==> Add a way to search N-Grams
- Multiple tabs
  - ==> Add multiple tabs within the application to allow the user to preserve old searches and settings.
- Visual Graphs
  - ==> Allow the user to make graphs in a new window to visualize queries.
- Concordance
  - ==> Add a setting to visually line up the titles by a keyword.
- Super Corpus
  - ==> Crawl the entire Prog Archives website for all entries within established criteria and make a "Super Corpus" with much more data to use.
- Text Click
  - ==> Allow the user to click a word in the explorer to see tagging info.



- Album Covers
  - ==> Allow the user to see covers associated with a selected album/track.
- Query Optimizations
  - ==> Optimize the search functionality to be faster.
- GUI Overhaul
  - ==> Make a more appealing/functional GUI.
- Add/Remove Entries
  - ==> Add more entries to the corpus, and remove problematic ones

## Generative Artificial Intelligence (Gen-AI) Usage Disclaimer

It can be assumed that Gen-AI was not used to generate neither prose, nor code related to this assignment with the following explicit exceptions in the file:

==> corpus\_explorer.py

Gen-AI tools (ChatGPT-4) were used for:

- The design of the search functionality (only for the layout of tag filtering statements)
- Code and ideas for the loading bar and parallel processing (worker functions, etc)
- Pyqt signals related to the functionality listed above
- and the unmodified save\_subset function (unused) for the unimplemented chain-search feature

Precisely no input from Gen-AI sources was used for any other files associated with this project.

## Sub-Genres Included

- Symphonic Prog
- Eclectic Prog
- Heavy Prog
- Psychedelic/Space Rock
- Progressive Electronic
- Prog Folk
- Jazz Rock/Fusion
- Canterbury Scene
- Rock Progressivo Italiano
- Krautrock
- Zeuhl
- RIO/Avant-Prog
- Crossover Prog
- Proto-Prog
- Neo-Prog
- Prog Related

# USAGE GUIDE

## First Row

- **Search Bar Text Field**

This is where you type the word to search for in the corpus. Multiple words are not supported.

- **Stats Button**

This button changes the View Window to show word frequencies in a "Corpus Statistics" page.

- **Part of Speech (POS) Dropdown**

This dropdown lets you select the POS tag to be searched that is associated with the word entered in the Search Bar. The tagset used is the NLTK tagset, which is a modified Penn tagset.

## Second Row

- **Search Button**

This button starts the search.

- **Clear Button**

This button clears the Search Box and the View Window

- **Case Checkbox**

This checkbox toggles Case-Sensitive searching. Checked means that the the search is Case-Sensitive

- **Lemma Search Checkbox**

This checkbox toggles the Lemma Search functionality. Checked means that only lemmas are searched in the corpus. You must put a lower-case lemma in the Search Bar to use this feature.

- **Word Index Checkbox**

This checkbox toggles the Word Index Search. It tells the explorer to only return results that match the Searched word at the specific number typed in the Word Index Text Field. Ex. In the sentence "The fox is hungry," the word "fox" is at index 2, and the word "hungry" is at index 4.

- **Word Index Text Field**

This text field is where you enter the index for the Word Index Search.

## Dates Tags

- **"All" Button**

This button checks all date checkboxes

- **"None" Button**

This button unchecks all date checkboxes

- **Date Checkboxes**

These checkboxes allow you to search by the album release date.

## Artist Tags

- **Artist Name Search**

The checkbox toggles the search using the text field to the right. This line is for searching by Artist name (Not Case-Sensitive).

- **Artist Sub-Genre Search**

The checkbox toggles the search using the dropdown to the right. This line is for searching by Sub-Genre (As appears on Prog Archives).

- **Artist Language Search**

The checkbox toggles the search using the dropdown to the right. This line is for searching by language. In many cases the group may have a different native language than is listed within the corpus. The language option indicates the primary language used for the album, even if it isn't the native language of the artist.

## Album Tags

- **Album Name Search**

The checkbox toggles the search using the text field to the right. This line is for searching by Album name (Not Case-Sensitive).

- **Album Rating Search**

The checkbox toggles the search using the text field to the right. The operator dropdown allows searching using the symbols ">", "=", "<" whereby the ">" returns results greater than the input, and the "<" returns results less than the input. This line is for searching by Album rating (As appears on Prog Archives) in the form "X.XX" where each "X" is a digit from 0-9. The scale used only goes from 1.00 to 5.00, and most results lie in the 3.XX range or the 4.XX range.

- **Album Length Search**

The checkbox toggles the search using the text field to the right. The operator dropdown allows searching using the symbols ">", "=", "<" whereby the ">" returns results greater than the input, and the "<" returns results less than the input. This line is for searching by Album length either in the form of seconds (integer) or in the styles "MM:SS", or "H:MM:SS" where H is the number of hours, MM is the number of minutes (2 digits), and SS is the number of seconds (2 digits).

## Track Tags

- **Track Index Search**

The checkbox toggles the search using the text field to the right. The operator dropdown allows searching using the symbols ">", "=", "<" whereby the ">" returns results greater than the input, and

the "<" returns results less than the input. This line is for searching by a numeric Track index (integer). Tracks originally listed with positions "a-g" or "i-vii" are converted to numeric form automatically.

- **Track Length Search**

The checkbox toggles the search using the text field to the right. The operator dropdown allows searching using the symbols ">", "=", "<" whereby the ">" returns results greater than the input, and the "<" returns results less than the input. This line is for searching by Track length either in the form of seconds (integer) or in the style "MM:SS" where MM is the number of minutes (2 digits), and SS is the number of seconds (2 digits).

- **Word Count Search**

The checkbox toggles the search using the text field to the right. The operator dropdown allows searching using the symbols ">", "=", "<" whereby the ">" returns results greater than the input, and the "<" returns results less than the input. This line is for searching by Track Word Count (integer). It tells the corpus to only return tracks that have the amount of words that satisfies the input. Some special punctuation cases are considered words (such as double-quotes).

- **Exclude Sub-Tracks**

When checked, filters out all sub-tracks in the corpus. This can be usefull when looking for tracks that are less than a given time, since sub-tracks have a length of zero and will show up, skewing results.

## View Window

Upon a normal search, the following information may appear:

- **"Hits"**

Indicates how many results (tracks) were matched.

- **"Lemma Frequency"**

Indicates the frequency that the lemma of the searched word appears.

- **"Title Frequency"**

Indicates the number of titles (hits) that contain the searched word.

- **Corpus Entry**

**The entries should appear in this format:**

```
- = - Group Name - = -
```

```
Track **Title** (Album Name)
```

```
=> *Original Text* If Translated
```

# ARTIST LIST

## The "Big Seven"

### Pink Floyd [en] (12)

- [The Piper at the Gates of Dawn](#)
- [A Saucerful of Secrets](#)
- [More](#)
- [Ummagumma](#)
- [Atom Heart Mother](#)
- [Meddle](#)
- [Obscured by Clouds](#)
- [The Dark Side of the Moon](#)
- [Wish You Were Here](#)
- [Animals](#)
- [The Wall](#)
- [The Final Cut](#)

### Genesis [en] (12)

- [From Genesis to Revelation](#)
- [Trespass](#)
- [Nursery Cryme](#)
- [Foxtrot](#)
- [Selling enland by the Pound](#)
- [The Lamb Lies Down on Broadway](#)
- [A Trick of the Tail](#)
- [Wind & Wuthering](#)
- [...And Then There Were Three...](#)
- [Duke](#)
- [Abacab](#)
- [Genesis](#)

### Yes [en] (11)

- [Yes](#)
- [Time and a Word](#)
- [The Yes Album](#)
- [Fragile](#)
- [Close to the Edge](#)
- [Tales from Topographic Oceans](#)
- [Relayer](#)
- [Going for the One](#)
- [Tormato](#)
- [Drama](#)
- [90125](#)

### Rush [en] (9)

- [Rush](#)
- [Fly by Night](#)
- [Caress of Steel](#)
- [2112](#)
- [A Farewell to Kings](#)
- [Hemispheres](#)
- [Permanent Waves](#)
- [Moving Pictures](#)
- [Signals](#)

#### King Crimson [en] (10)

- [In the Court of the Crimson King](#)
- [In the Wake of Poseidon](#)
- [Lizard](#)
- [Islands](#)
- [Larks' Tongues in Aspic](#)
- [Starless and Bible Black](#)
- [Red](#)
- [Discipline](#)
- [Beat](#)
- [Three of a Perfect Pair](#)

#### Jethro Tull [en] (14)

- [This Was](#)
- [Stand Up](#)
- [Benefit](#)
- [Aqualung](#)
- [Thick as a Brick](#)
- [A Passion Play](#)
- [War Child](#)
- [Minstrl in the Gallery](#)
- [Too Old To Rock 'n' Roll Too Young to Die!](#)
- [Songs from the Wood](#)
- [Heavy Horses](#)
- [Stormwatch](#)
- [A](#)
- [The Broadsword and the Beast](#)

#### Emerson Lake and Palmer [en] (7)

- [Emerson Lake & Palmer](#)
- [Pictures at an Exhibition](#)
- [Tarkus](#)
- [Trilogy](#)
- [Brain Salad Surgery](#)
- [Works Vol. 1](#)

- [Works Vol. 2](#)

## Secondary Bands

### Gentle Giant [en] (11)

- [Gentle Giant](#)
- [Aquiring the Taste](#)
- [Three Friends](#)
- [Octopus](#)
- [In a Glass House](#)
- [The Power and the Glory](#)
- [Free Hand](#)
- [Interview](#)
- [The Missing Piece](#)
- [Giant for a Day](#)
- [Civilian](#)

### Van Der Graaf Generator [en] (8)

- [The Aerosol Grey Machine](#)
- [The Least We Can Do Is Wave to Each Other](#)
- [H to He, Who am the Only One](#)
- [Pawn Hearts](#)
- [Godbluff](#)
- [Still Life](#)
- [World Record](#)
- [The Quiet Zone - The Pleasure Dome](#)

### Camel [en] (10)

- [Camel](#)
- [Mirage](#)
- [The Snow Goose](#)
- [Moonmadness](#)
- [Rain Dances](#)
- [Breathless](#)
- [I Can See Your House From Here](#)
- [Nude](#)
- [The Single Factor](#)
- [Stationary Traveller](#)

### Alan Parsons Project [en] (7)

- [Tales of Mystery and Imagination](#)
- [I Robot](#)
- [Pyramid](#)
- [Eve](#)
- [The Turn of a Friendly Card](#)

- [Eye in the Sky](#)
- [Ammonia Avenue](#)

#### Renaissance [en] (9)

- [Renaissance](#)
- [Illusion](#)
- [Prologue](#)
- [Ashes Are Burning](#)
- [Turn of the Cards](#)
- [Scheherazade and Other Stories](#)
- [Novella](#)
- [A Song for All Seasons](#)
- [Azure d'Or](#)

#### Eloy [en] (10)

- [Eloy](#)
- [Inside](#)
- [Floating](#)
- [Power and the Passion](#)
- [Dawn](#)
- [Ocean](#)
- [Silent Cries and Mighty Echoes](#)
- [Colours](#)
- [Planets](#)
- [Time to Turn](#)

#### Island [en] (1)

- [Pictures](#)

#### Focus [en] (6)

- [In and Out of Focus](#)
- [Moving Waves](#)
- [Focus 3](#)
- [Hamburger Concerto](#)
- [Mother Focus](#)
- [Ship of Memories](#)

#### Wishbone Ash [en] (5)

- [Wishbone Ash](#)
- [Pilgrimage](#)
- [Argus](#)
- [Wishbone Four](#)
- [There's the Rub](#)

#### Curved Air [en] (4)



- [Airconditioning](#)
- [Second Album](#)
- [Phantasmagoria](#)
- [Air Cut](#)

#### Tangerine Dream [en] (3)

- [Phaedra](#)
- [Rubycon](#)
- [Stratosfear](#)

#### Hawkwind [en] (5)

- [Hawkwind](#)
- [X in Search of Space](#)
- [Doremi Fasol Latido](#)
- [Hall of the Mountian Grill](#)
- [Warrior on the Edge of Time](#)

#### Nektar [en] (6)

- [Journey to the Center of the Eye](#)
- [A Tab in the Ocean](#)
- [...Sounds Like This](#)
- [Remember the Future](#)
- [Down to Earth](#)
- [Recycled](#)

#### Triumvirat [en] (5)

- [Mediterranean Tales \(Across the Waters\)](#)
- [Illusions on a Double Dimple](#)
- [Spartacus](#)
- [Old Loves Die Hard](#)
- [Pompeii](#)

#### Ekseption [en] (8)

- [Ekseption](#)
- [Beggar Julia's Time Trip](#)
- [Ekseption 3](#)
- [00-04](#)
- [Ekspetion 5](#)
- [Trinity](#)
- [Bingo](#)
- [Mindmirror](#)

#### Trace [en] (2)

- [Trace](#)
- [Birds](#)

## UK [en] (2)

- [UK](#)
- [Danger Money](#)

## Supertramp [en] (4)

- [Crime of the Century](#)
- [Crisis What Crisis](#)
- [Even in the Quietest Moments...](#)
- [Breakfast in America](#)

## Strawbs [en] ()

- [Hero and Heroine](#)
- [Ghosts](#)

## Led Zeppelin [en] (8)

- [Led Zeppelin](#)
- [Led Zeppelin II](#)
- [Led Zeppelin III](#)
- [Led Zeppelin IV](#)
- [Houses of the Holy](#)
- [Physical Graffiti](#)
- [Presence](#)
- [In Through the Out Door](#)

## The Who [en] (3)

- [Tommy](#)
- [Who's Next](#)
- [Quadrophenia](#)

## Deep Purple [en] (11)

- [Shades of Deep Purple](#)
- [The Book of Taliesyn](#)
- [Deep Purple](#)
- [Deep Purple in Rock](#)
- [Fireball](#)
- [Machine Head](#)
- [Who Do We Think We Are](#)
- [Burn](#)
- [Stormbringer](#)
- [Come Taste the Band](#)
- [Perfect Strangers](#)

## Uriah Heep [en] (6)

- [Uriah Heep](#)

- [Salisbury](#)
- [Look At Yourself](#)
- [Demons and Wizards](#)
- [The Magician's Birthday](#)
- [Sweet Freedom](#)

The Moody Blues [en] (1)

- [Days of Future Passed](#)

Kansas [en] (7)

- [Kansas](#)
- [Song for America](#)
- [Masque](#)
- [Leftoverture](#)
- [Point of Know Return](#)
- [Monolith](#)
- [Audio Visions](#)

Styx [en] (11)

- [Styx](#)
- [Styx II](#)
- [The Serpenth Is Rising](#)
- [Man of Miracles](#)
- [Equinox](#)
- [Crystal Ball](#)
- [The Grand Illusion](#)
- [Pieces of Eight](#)
- [Cornerstone](#)
- [Paradise Theatre](#)
- [Kilroy Was Here](#)

Solo Artists

Frank Zappa [en] (31)

- [Freak Out!](#)
- [Absolutely Free](#)
- [Lumpy Gravy](#)
- [We're Only in It for the Money](#)
- [Uncle Meat](#)
- [Hot Rats](#)
- [Burnt Weeny Sandwich](#)
- [Weasels Ripped My Flesh](#)
- [Chunga's Revenge](#)
- [200 Motels](#)
- [Waka - Jawaka](#)

- [The Grand Wazoo](#)
- [Over-Nite Sensation](#)
- [Apostrophe \('\)](#)
- [Roxy & Elsewhere](#)
- [One Size Fits All](#)
- [Bongo Fury](#)
- [Zoot Allures](#)
- [Zappa in New York](#)
- [Studio Tan](#)
- [Sleep Dirt](#)
- [Sheik Yerbouti](#)
- [Orchestral Favorites](#)
- [Joe's Garage Act I](#)
- [Joe's Garage Acts II & III](#)
- [Tinsel Town Rebellion](#)
- [Shut Up 'N Play Yer Guitar](#)
- [Shut Up 'N Play Yer Guitar Some More](#)
- [Return Of The Son Of Shut Up 'N Play Yer Guitar](#)
- [You Are What You Is](#)
- [Ship Arriving Too Late to Save a Drowning Witch](#)

Steve Hackett [en] (4)

- [Voyage of the Acolyte](#)
- [Please Don't Touch!](#)
- [Spectral Mornings](#)
- [Defector](#)

Peter Gabriel [en] (4)

- [Car](#)
- [Scratch](#)
- [Melt](#)
- [Security](#)

Michael Rutherford [en] (1)

- [Smallcreep's Day](#)

Anthony Philips [en] (7)

- [The Geese and the Ghost](#)
- [Wise After the Event](#)
- [Private Parts & Pieces](#)
- [Sides](#)
- [Private Parts & Pieces II - Back to the Pavillion](#)
- [1984](#)
- [Private Parts & Pieces III - Antiques](#)

Chris Squire [en] (1)

- [Fish Out of Water](#)

Rick Wakeman [en] (1)

- [The Six Wives of Henry VIII](#)

Bill Bruford [en] (2)

- [Feels Good to Me](#)
- [One of a Kind](#)

Peter Hammill [en] (6)

- [Fool's Mate](#)
- [Chameleon in the Shadow of the Night](#)
- [The Silent Corner and the Empty Stage](#)
- [In Camera](#)
- [Nadir's Big Chance](#)
- [Over](#)

Mike Oldfield [en] (8)

- [Tubular Bells](#)
- [Hergest Ridge](#)
- [Ommadawn](#)
- [Incantations](#)
- [Platinum](#)
- [QE2](#)
- [Five Miles Out](#)
- [Crises](#)

Klaus Schultz [en] (3)

- [Timewind](#)
- [Moondawn](#)
- [Mirage](#)

Canterbury Scene

Caravan [en] (7)

- [Caravan](#)
- [If I Could Do It All Over Again, I'd Do It All Over You](#)
- [In the Land of Grey and Pink](#)
- [Waterloo Lily](#)
- [For Girls Who Grow Plump in the Night](#)
- [Cunning Stunts](#)
- [Blind Dog at St. Dunstons](#)

Soft Machine [en] (9)

- [The Soft Machine](#)

- [Volume Two](#)
- [Third](#)
- [Fourth](#)
- [Fifth](#)
- [Six](#)
- [Seven](#)
- [Bundles](#)
- [Softs](#)

Matching Mole [en] (2)

- [Maching Mole](#)
- [Little Red Record](#)

Robert Wyatt [en] (1)

- [Rock Bottom](#)

Hatfield and the North [en] (2)

- [Hatfield and the North](#)
- [The Rotters' Club](#)

National Health [en] (2)

- [National Health](#)
- [Of Queues and Cures](#)

Quiet Sun [en] (1)

- [Mainstream](#)

Gryphon [en] (5)

- [Gryphon](#)
- [Midnight Mushrumps](#)
- [Red Queen to Gryphon Three](#)
- [Raindance](#)
- [Treason](#)

Egg [en] (3)

- [Egg](#)
- [The Polite Force](#)
- [The Civil Surface](#)

Khan [en] (1)

- [Space Shanty](#)

Gong [en] (8)

- [Magick Brother](#)

- [Camembert Electrique](#)
- [Radio Gnome Invisible Part 1 - Flying Teapot](#)
- [Radio Gnome Invisible Part 2 - Angel's Egg](#)
- [Radio Gnome Invisible Part 3 - You](#)
- [Shamal](#)
- [Gazeuse!](#)
- [Espresso II](#)

Steve Hillage [en] (1)

- [Fish Rising](#)

Magma [en] (7)

- [Magma](#)
- [1001 Centigrades](#)
- [Mekanik Destruktiv Kommandoh](#)
- [Qurdah Itah](#)
- [Kohntarkosz](#)
- [Udu Wudu](#)
- [Attahk](#)

Fusion

Mahavishnu Orchestra [en] (4)

- [The Inner Mounting Flame](#)
- [Birds of Fire](#)
- [Apocalypse](#)
- [Visions of the Emerald Beyond](#)

Return to Forever [en] (1)

- [Romantic Warrior](#)

Brand X [en] (3)

- [Unorthodox Behaviour](#)
- [Moroccan Roll](#)
- [Masques](#)

Arti E Mestieri [it] (1)

- [Tilt - Immagini Per Un Orecchio](#)

Spin [en] (2)

- [Spin](#)
- [Whirlwind](#)

Krautrock

Can [en] (3)

- [Tago Mago](#)
- [Ege Bamyasi](#)
- [Future Days](#)

Amon Duul II [en] (1)

- [Yeti](#)

Neu! [de] (1)

- [Neu!](#)

Rock Progressivo Italiano (RPI)

Premiata Forneria Marconi [it] (Award-Winning Marconi Bakery) (3)

- [Storia Di Un Minuto](#)
- [Per Un Amico](#)
- [L'Isola Di Niente](#)

Banco Del Mutuo Soccorso [it] (Bank of Mutual Trust) (3)

- [Banco Del Mutuo Soccorso](#)
- [Darwin!](#)
- [Io Sono Nato Libero](#)

Le Orme [it] (The Footsteps) (4)

- [Collage](#)
- [Uomo Di Pezza](#)
- [Felona E Sorona](#)
- [Contrappunti](#)

Area [it] (4)

- [Arbeit Macht Frei](#)
- [Caution Radiation Area](#)
- [Crac!](#)
- [Meledetti](#)

Franco Battiato [it] (4)

- [Fetus](#)
- [Pollution](#)
- [Sulle Corde Di Aries](#)
- [Clic](#)

Quella Vecchia Locanda [it] (That Old Inn) (2)

- [Quella Vecchia Locanda](#)
- [Il Tempo della Gioia](#)



Museo Rosenbach [it] (Rosenbach Museum) (1)

- [Zarathustra](#)

Il Balletto Di Bronzo [it] (The Bronze Ballet) (1)

- [Ys](#)

Il Paesi Di Ballochi [it] (The Land of Toys) (1)

- [Il Paese Dei Balocchi](#)

L'Uovo Di Columbo [it] (The Egg of Columbus) (1)

- [L'Uovo Di Columbo](#)

Biglietto Per L'Inferno [it] (Ticket to Hell) (1)

- [Biglietto Per L'Inferno](#)

New Trolls [it] (4)

- [Concerto Grosso Per I New Trolls](#)
- [Searching for a Land](#)
- [UT](#)
- [Concerto Grosso No 2](#)

Formula 3 [it] (1)

- [Sognando e Risognando](#)

Il Volo [it] (2)

- [Il Volo](#)
- [Essere O Non Essere ?](#)

Maxophone [it] (1)

- [Maxophone](#)

Alphataurus [it] (1)

- [Alphataurus](#)

Osana [it] (2)

- [L'Uomo](#)
- [Palepoli](#)

Blocco Mentale [it] (Mental Block) (1)

- [POA](#)

Raccomandata con Ricevuta di Ritorno [it] (Registered Mail with Return Receipt) (1)

- [Per... Un Mondo Di Cristallo](#)

Picchio Dal Pozzo [it] (Woodpecker from the Well) (1)

- [Picchio Dal Pozzo](#)

Semiramis [it] (1)

- [Dedicato A Frazz](#)

Celeste [it] (1)

- [Principe Di Un Giorno](#)

Metamorfosi [it] (Metamorphosis) (1)

- [Inferno](#)

Locanda Delle Fate [it] (Fairy Inn) (1)

- [Forse Le Lucciole Non Si Amano Più](#)

Il Rovescio Della Medaglia [it] (The Reverse of a Coin) (1)

- [Contaminazione](#)

Reale Accademia Di Musica [it] (Royal Academy of Music) (1)

- [Reale Accademia Di Musica](#)

Apoteosi [it] (Apotheosis) (1)

- [Apoteosi](#)

Campo Di Marte [it] (Field of Mars) (1)

- [Campo Di Marte](#)

Murple [it] (1)

- [Io Sono Murple](#)

International

Bubu [es] (1)

- [Anabelas](#)

Crucis [es] (Cross) (2)

- [Crucis](#)
- [Los Delirios Del Mariscal](#)

Fusioon [es] (3)

- [Fusioon](#)

- [Fusioon 2](#)
- [Minorisa](#)

Harmonium [fr] (1)

- [Si On Avait Besoin D'une Cinquième Saison](#)

Avtograf [ru] (Autograph) (2)

- [Avtograf 1](#) <-- Add cyrillic manually
- [Avtograf 2](#) <-- YouTube Link (Add Manually)

Horizont [ru] (Horizon) (1)

- [Summer in Town](#)

East [hu] (2)

- [Jatekok](#)
- [Huseg](#)

Artists To Include:

- The Who ✓
- Frank Zappa ✓  
Shut up 'n play yer guitar 1, 2, 3 (NOT THE BOX SET)
- Deep Purple ✓
- Uriah Heep ✓
- Horizont ✓

TAGGING

Artist Tags

artist-name  
Name of the artist.

sub-genre  
Prog archives sub-genre.

language  
Main language of the artist

Album Tags

album-name  
- Name of the album.

album-rating  
- Prog archives cumulative user rating.

```
album-length
 - Duration of the album.
```

## Track Tags

```
tr-index
 - The total index of the track in the album. Do not restart count when the
 side changes.
tr-lenth
 - Duration of the track.
w-count
 - Word count of the track.
*?tr-type (could be automated, NOT IMPLEMENTED)
 - standard
 - main-track
 - part-track
 - compound-track
```

## Word Tags

```
pos
 - Part of Speech of the word (using an auto-tagger).
lemma
 - The lemmatized form of the word (ie. Dogs, doggies, dog, dawg ->
 lemma="dog").
```