Chord progression pipeline – choices

1. Data sources:
   1. Amateur chord sites (crawling + parsing)
   2. Midi song databases (crawling + chord recognition)
   3. Manually annotated songs (just read the files)
   4. Audio -> midi -> chords (come on…)
2. Chord model:
   1. Modulus –
      1. By starting tonic (only from 1.c or our processing)
         * Separates same pattern in different tonics
      2. By current tonic (only processing, quite unfeasible)
         * Aggregates same progression in different tonics
      3. By running differences (easy processing)
         * Feasible
         * Aggregates same progressions in different tonics
      4. Combination??
   2. Chord vectors to frequency vectors:
      1. Unsupervised - find out which progressions appear frequently. Combinatorial - go over variable length frames in chords vectors, and find sufficiently frequent patterns. Then, go over each song again and compute the occurrences of patterns in the song. Variations:
         * Window lengths, occurrence threshold are parameters to be chosen.
         * Differentiate between patterns that start on different beats, measures, etc.
         * Differentiate between patters with variable duration of chords.
         * Normalize appearances by song/progression length and/or progressions that overlap themselves
           1. When finding out which progressions are frequent in the data-set, normalize by progression length alone, and don't count overlaps (actively remove duplicate counts).
           2. When computing actual frequency vectors for songs, normalize by song length and progression length, and don't count overlaps(actively remove duplicate counts).
3. Tagging data:
   1. Tag thresholding
      1. Certainty value
      2. Frequent tags
      3. Frequently tagged songs
   2. Tag redundancy
      1. similar strings for same shit
         * we can apply string similarity algorithms to union tags named similarly (‘rock’, ’Rock’)
      2. different genres for same song
4. Research questions:
   1. Unknown frequent harmonic progressions
   2. Our own tagging system – unsupervised frequency vectors clustering. Comparison with known tags.
   3. Tag phylogeneza
   4. Composing assistance
      * By recommending songs (chord vectors) matching given tags
      * By (somehow) generating a chord vector/recommended progressions matching given tags.
   5. Song recommendation system, by simple distance function over frequency vectors.