0.1 Mission Statement

This is our personal collection of some important ideas in computer science and programming. It shall help us document high-level insights that can serve us in practice. We do not try to capture topics in-depth, but instead just try to make us aware, so that we know what to look for when we are in need. We are aware that when everything is important, nothing is important.

We also try to document some of our occasional nuggets of temporary insights.

0.2 Main Topics

- Nothing substantial yet

0.3 Sources to review

- Study notes
  - Compiler Construction
  - Experimental Economics: Design Elements (e.g., basline neighborhood) and the basic idea of falsification as supplement material for the algorithm engineering notes
  - Formal Systems (something in there?)
  - ~~Advanced Data Structures~~
  - Parallel Algorithms
  - Parallel Machines and Parallel Programming
  - Randomized Algorithms
  - Computerarchitecture (implications of branch predictions, cache coherece, pipelining, super scalar architecutures . . .)
  - Softwareengineering II
  - Game Theory (Battle of the sexes :p)

- Lecture notes and books to skim
  - Algorithm Engineering
  - Algorithm II
  - Algorithms and Data Structures — The Basic Toolbox
  - Linder's *Things Thy Should have taught you*

- Books with potential:
  - Pragmatic Thinking and Learning
  - Notes from the Pragmatic Programmer

– Notes from Head First Software Development (Very light stuff but maybe there is something in there. I used to enjoy reading it.)

– . . .

1 Desigining Abstractions

- Single Level of Abstraction
- Single Responsibility Principle
- Separation of Concerns
- Interface Segragation Principle
- Liskov Substitution Principle
- Open Closed Principle
- Law of Demeter

If I am not totally mistaken, these principles are well described within SICP.

2 Algorithm & Datastructures Design Toolbox

2.1 Data Structure Use & Abuse

2.1.1 Trie

A *trie* is a special kind of edge labeled tree. If used over a collection of keys, it can be used to find the significant difference between these keys (i.e., the *important bit positions* sufficient to distinguish all keys). For sorted sequences it can be constructed in $\mathcal{O}(n)$. Examples:

- *Fusion Trees* compress keys within *B-Tree* nodes by reducing them to the bits at the important bit positions. All keys of a node can then be fused into a single machine word and compared to a query element in parallel using a single bitparallel subtraction.

- *String B-Trees* use *Patricia Tries* to identify the subset of characters relevant for the comparison of a string pattern to the keys of a *B-Tree*. Knowing these characters and the lengths of the common prefixes of keys, the number of required I/Os per traversal can be limited.

- *Signature Sort* splits keys into chunks and compresses these chunks using a hash function. A trie is then used to filter the chunks that are not relevant for the sort order of the keys. The keys therefore become smaller and sorting them becomes easier.

2.2   Recurring Design Ideas

There seem to be some simple ideas that have influenced the desgin of many algorithms and datastructures. This list is not authoritative (in particular w.r.t. to naming), it just lists some observations. The ideas seem to be heavily connected and seldomly used in isolation.

- **Tradeoff:** Inspect simple, naive solutions with contradicting space and runtime bounds (e.g, online computation vs. precomputation of all results) and try to find a tradeoff that achieves the best of both worlds. So for example, instead of precoumputing all results, precompute just a particular subset. Examples:

  - Level Ancestor Queries (to find the tree-ancestor on a particular level) can be approached using *jump pointers* and *ladder decomposition*, which both precompute a subset of all potential queries and run in $\mathcal{O}(\log n)$ time. However, a combination of both leads to constant query time and a linear space requirement, thus heavily improving over the computation of all results.

- **Combination:** Combine several different datastructures so that expensive operations of one datastructure can be improved using a specialized data structure for this particular sub-problem. Examples:

  - *String B-Trees* build *Patricia Tries* on top of individual B-Tree nodes to improve over the binary search to find the insertion point within the keys of a node.

- **Indirection/Bucketing:** Instead of working on a large problem set (with larger accompanying space & runtime bounds), first find a suitable bucket and then solve the problem only within this particular bucket. Examples:

  - *Perfect Hashing* uses an indirection so that it is not required to find a perfect (injective) hash function for all elemenets, but only for the elements within the same collision bucket. This reduces the overall space requirement while retaining the expected construction and query time.
  - *Y-Fast Tries* improve upon the space requirements of *X-Fast Tries* by building the trie over representatives of buckets instead of over all elements. The buckets are implemented as balanced binary trees and do not just reduce the space requirement of *X-Fast Tries* but also help to reduce update costs using amortization.
  - *Succinct Data Structures* can profit from indirection layers as they reduce the number of elements that have to be distinguished from the whole universe down to all elements within a bucket. Thus, to dinstinguish these elements less bits are required.

- **Decomposition:** Instead of dealing with elements of objects as a whole, split them into logical subparts. Exploit this inner structure to achieve higher efficiency and eventually solve the problem within a richer model of computation (e.g., compare bits instead of whole integers) . Examples:
  - *X-Fast Tries* split integer keys into their common prefixes and use a binary search to find the longest common prefix.
  - *van Emde Boas Trees* split recursively split keys into a top and bottom halve in order to efficiently navigate within buckets and summaries over these buckets.
  - Level Ancestor Queries can be implemented via the idea of recursively splitting a tree into its longests paths (*longest path decomposition*) and precomputing the results within these paths.

- **Input Reduction/Simplification:** Reduce the problem set at hand until it is simple enough so that it can be solved using basic means. This idea is at the core of most recursive algorithms.
  - *Independet-Set Removal* is used to speed up list ranking.
  - *Signature Sort* recursively reduces the size of the keys until they can be sorted with *Packed Sorting*.
  - The *findclose* operation on *Succinct Trees* finds the position of a closing bracket within the bitstring representing the tree. The idea is to solve the problem for a corresponding, so-called pioneer and then deduce the actual closing position from the closing position of the corresponding pioneer.

- **Broadword Computing:** Fuse data elements into machine words and run operation on then in parallel, instead of sequentially. You can get more done within fewer instructions and without having to hit memory.
  - *Fusion Trees* compare several integer keys with a single bitparallel computation.
  - *Packed Sorting* is a variant of mergesort that packs several short integer keys into a machine word. It then uses bitparallel computations to speed up the base case and the merge of sorted words. In particular, it relies on a bitparallel version of *Bitonic Sorting*.

2.3 Succinct Data Structures

Succinct Data Structures are space efficient implementations of abstract data types (e.g., trees, bit sets) that still allow for efficient queries.

2.4 Randomized Algorithms

- Among others, randomized algorithms help to improve robustness concerning worst-case inputs (e.g., think of the pivot selection problem in `quicksort` when the sequence is provided by a malicious adversary).

- Allowing randomized algorithms to compute a *wrong* result (with a very low probability) can open many new possibilities concerning speed, space, quality and ease of implementation (e.g., think of *bloom filters* or *Approximate Distance Oracles*).

- Never use the `C rand()` function. If in doubt, use *Mersenne Twister*.

- Use random numbers with care. Treat them as a scarce resource.

- In certain parallel setups, *expected* bounds of randomized algorithms do no longer hold: Consider `n` processes that call operations with *expected* runtime bounds and that have to be synchronized before and afterwards. The runtime will suffer whenever at least one of the processes hits an expensive case.

3 Personal Development

- Always life on the edge of incompetence. When ever you have the chance, approach challanges beyond your current skill level. Be focused, determined and passioned and you will be invincible, because even if you fail (what you won't), you have learned a great deal.

- There is no speed limit in life. You don't have to move (improve, learn, . . .) as slowly and overcautiously as others.

- Never presume to already know. Have a beginner's mind.

- Awareness trumps trying.

- Act or accept there is nothing in between.

- "Money is nothing more than neutral proof that you're adding value to people's lives. Making sure you're making money is just a way of making sure you're doing something of value to others." (Derek Sivers)

3.1 Studying

- Approach study tasks like you approach programming taks (see the common principles)

- Do not expect teachers to teach you (they only present information). Nobody will teach you anything. You have to teach yourself.

- Studying is not about time, it's about effort. You have to approach a difficult topics from different angles and discuss it with peers. Make the ideas your own by discovering them for yourself (e.g., with your own examples). Stupid repetition just does not cut it. If you have to memorize something, you haven't understood it.

- Don't miss the big picture by getting lost in all the details. Always zoom-out again and fish for connections and insights.

- Always write down key insights as succinctly and precicely as possible. Feel the pain. You have to do it, because it is a different thing to believe to understand something and to be able to articulate it.

- If stuck with anything, ask yourself *what exactly_* is it, that you don't understand? (See the notes on debugging).

3.2 Stress & Burnout Prehab

- Stress kills creativity and gives you a bad hangover. Don't get back to work just because you are bored. Once your creativity returns (with ideas on cool projects, the strong desire to read one the countless unread books on your shelf, people you really want to meet again, . . .) you will feel more energetic than ever and can re-approach your life (and work) with a whole new mind.

3.3 Mentoring & Teaching

- Challenge the student by constantly asking questions. Let the student prove they got it. Set a high pace.

4 Sources

- There's no speed limit and 6 things I wish I knew the day I started Berklee by Derek Sivers

5 Programming

5.1 Common Principles

- Think hard before you start.

- Focus. Eliminate distractions and obstacles. If you cannot stay within *the flow*, you are doing something wrong.

- *Make it*, *make it run* (so that the tests pass), *make it better* (review and refactor until you can honestly say that it is good enough)

- If in doubt, leave it out (abstractions, features, . . .), because *you ain't gonna need it (YAGNI)!* You can always come back later. Don't waste your time now.

5.2   Object-oriented Programming

- Using an OO-language does not imply you *'got it'* and are doing it right.

- If nothing else, do atleast try to follow the *tell, don't ask* principle. ## Testing

- Look for *inflection points* as the sweet spot between testing effort and covered functionality.

5.3   Debugging

- Again, think before you start. Relax, breath and fetch your favorite beverage. Find the bug in your head. Only then start the debugger, if at all!

- Debugging is not a reason to stay overtime. Document your thought process so that you know where to pick up the next day.

- If stuck, describe your problem to an (imaginary) colleague.

- Dig until you have found the root cause. If you are inclined to just add a `null` check, you haven't found it yet.