Indexes

• *Indexes* are data structures designed to make search faster
• Text search has unique requirements, which leads to unique data structures
• Most common data structure is *inverted index*
  – general name for a class of structures
  – “inverted” because documents are associated with words, rather than words with documents
  • similar to a *concordance*
Indexes and Ranking

• Indexes are designed to support search
  – faster response time, supports updates

• Text search engines use a particular form of search: ranking
  – documents are retrieved in sorted order according to a score computing using the document representation, the query, and a ranking algorithm

• What is a reasonable abstract model for ranking?
  – enables discussion of indexes without details of retrieval model
Abstract Model of Ranking

Fred's Tropical Fish Shop is the best place to find tropical fish at low, low prices. Whether you're looking for a little fish or a big fish, we've got what you need. We even have fake seaweed for your fishtank (and little surfboards too).
More Concrete Model

\[ R(Q, D) = \sum_i g_i(Q) f_i(D) \]

- \( f_i \) is a document feature function
- \( g_i \) is a query feature function

Fred's Tropical Fish Shop is the best place to find tropical fish at low, low prices. Whether you're looking for a little fish or a big fish, we've got what you need. We even have fake seaweed for your fishtank (and little surfboards too).
Inverted Index

• Each index term is associated with an *inverted list*
  – Contains lists of documents, or lists of word occurrences in documents, and other information
  – Each entry is called a *posting*
  – The part of the posting that refers to a specific document or location is called a *pointer*
  – Each document in the collection is given a unique number
  – Lists are usually *document-ordered* (sorted by document number)
Example “Collection”

\[ S_1 \] Tropical fish include fish found in tropical environments around the world, including both freshwater and salt water species.

\[ S_2 \] Fishkeepers often use the term tropical fish to refer only those requiring fresh water, with saltwater tropical fish referred to as marine fish.

\[ S_3 \] Tropical fish are popular aquarium fish, due to their often bright coloration.

\[ S_4 \] In freshwater fish, this coloration typically derives from iridescence, while salt water fish are generally pigmented.

Four sentences from the Wikipedia entry for *tropical fish*
Simple Inverted Index

and 1
aquarium 3
are 3 4
around 1
as 2
both 1
bright 3
coloration 3 4
derives 4
due 3
environments 1
fish 1 2 3 4
fishkeepers 2
found 1
fresh 2
freshwater 1 4
from 4
generally 4
in 1 4
include 1
including 1
iridescence 4
marine 2
often 2 3
only 2
pigmented 4
popular 3
refer 2
referred 2
requiring 2
salt 1 4
saltwater 2
species 1
term 2
the 1 2
their 3
this 4
those 2
to 2 3
tropical 1 2 3
typically 4
use 2
water 1 2 4
while 4
with 2
world 1
Inverted Index with counts

- supports better ranking algorithms

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>aquarium</td>
<td>3:1</td>
</tr>
<tr>
<td>are</td>
<td>3:1</td>
</tr>
<tr>
<td>around</td>
<td>3:1</td>
</tr>
<tr>
<td>as</td>
<td>3:1</td>
</tr>
<tr>
<td>both</td>
<td>1:1</td>
</tr>
<tr>
<td>bright</td>
<td>3:1</td>
</tr>
<tr>
<td>coloration</td>
<td>3:1</td>
</tr>
<tr>
<td>derives</td>
<td>4:1</td>
</tr>
<tr>
<td>due</td>
<td>3:1</td>
</tr>
<tr>
<td>environments</td>
<td>1:1</td>
</tr>
<tr>
<td>fish</td>
<td>1:2</td>
</tr>
<tr>
<td>fishkeepers</td>
<td>1:1</td>
</tr>
<tr>
<td>found</td>
<td>2:1</td>
</tr>
<tr>
<td>fresh</td>
<td>2:1</td>
</tr>
<tr>
<td>freshwater</td>
<td>1:1</td>
</tr>
<tr>
<td>from</td>
<td>4:1</td>
</tr>
<tr>
<td>generally</td>
<td>4:1</td>
</tr>
<tr>
<td>in</td>
<td>1:1</td>
</tr>
<tr>
<td>include</td>
<td>1:1</td>
</tr>
<tr>
<td>including</td>
<td>1:1</td>
</tr>
<tr>
<td>iridescence</td>
<td>4:1</td>
</tr>
<tr>
<td>marine</td>
<td>2:1</td>
</tr>
<tr>
<td>often</td>
<td>2:1</td>
</tr>
</tbody>
</table>
## Inverted Index with positions

- supports proximity matches

<table>
<thead>
<tr>
<th>Term</th>
<th>Positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>and</td>
<td>1,15,3,5</td>
</tr>
<tr>
<td>aquarium</td>
<td>3,3,4,14</td>
</tr>
<tr>
<td>are</td>
<td>1,9,2,21</td>
</tr>
<tr>
<td>around</td>
<td>1,13,3,11</td>
</tr>
<tr>
<td>bright</td>
<td>3,12,4,5</td>
</tr>
<tr>
<td>coloration</td>
<td>4,7,3,7</td>
</tr>
<tr>
<td>derives</td>
<td>2,7,2,18</td>
</tr>
<tr>
<td>due</td>
<td>2,28,2,23</td>
</tr>
<tr>
<td>environments</td>
<td>3,2,3,18</td>
</tr>
<tr>
<td>fish</td>
<td>3,6,4,3</td>
</tr>
<tr>
<td>fishkeepers</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>found</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>fresh</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>freshwater</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>from</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>generally</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>in</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>include</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>including</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>iridescence</td>
<td>4,13,3,9</td>
</tr>
<tr>
<td>marine</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>often</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>only</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>pigmented</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>popular</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>refer</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>referred</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>requiring</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>salt</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>saltwater</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>species</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>term</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>their</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>this</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>those</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>to</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>tropical</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>typically</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>use</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>water</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>while</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>with</td>
<td>2,2,2,22</td>
</tr>
<tr>
<td>world</td>
<td>2,2,2,22</td>
</tr>
</tbody>
</table>
Proximity Matches

- Matching phrases or words within a window
  - e.g., "tropical fish", or “find tropical within 5 words of fish”

- Word positions in inverted lists make these types of query features efficient
  - e.g.,

<table>
<thead>
<tr>
<th>tropical</th>
<th>1,1</th>
<th>1,7</th>
<th>2,6</th>
<th>2,17</th>
<th>3,1</th>
</tr>
</thead>
<tbody>
<tr>
<td>fish</td>
<td>1,2</td>
<td>1,4</td>
<td>2,7</td>
<td>2,18</td>
<td>2,23</td>
</tr>
<tr>
<td></td>
<td>3,2</td>
<td>3,6</td>
<td>4,3</td>
<td>4,13</td>
<td></td>
</tr>
</tbody>
</table>

A better example would include where both "tropical" and "fish" occur, but are not adjacent.
Fields and Extents

• Document structure is useful in search
  – field restrictions
    • e.g., date, from:, etc.
  – some fields more important
    • e.g., title

• Options:
  – separate inverted lists for each field type
  – add information about fields to postings
  – use extent lists
Extent Lists

• An *extent* is a contiguous region of a document
  – represent extents using word positions
  – inverted list records all extents for a given field type
  – e.g.,

```
fish    1,2  1,4  2,7  2,18  2,23  3,2  3,6  4,3  4,13
  title  1:(1,3)  2:(1,5)  4:(9,15)
```
Other Issues

• Precomputed scores in inverted list
  – e.g., list for “fish” [(1:3.6), (3:2.2)], where 3.6 is total feature value for document 1
  – improves speed but reduces flexibility

• Score-ordered lists
  – query processing engine can focus only on the top part of each inverted list, where the highest-scoring documents are recorded
  – very efficient for single-word queries

note: proximity information is lost in a score-ordered list
Compression

• Inverted lists are very large
  – e.g., 25-50% of collection for TREC collections using Indri search engine
  – Much higher if n-grams are indexed

• Compression of indexes saves disk and/or memory space
  – Typically have to decompress lists to use them
  – Best compression techniques have good compression ratios and are easy to decompress

• Lossless compression – no information lost
Compression

• **Basic idea:** Common data elements use short codes while uncommon data elements use longer codes
  – Example: coding numbers
    • number sequence:
    • possible encoding:
      • encode 0 using a single 0:
        • only 10 bits, but...

0, 1, 0, 3, 0, 2, 0
00 01 00 10 00 11 00
14 bits

0 01 0 10 0 11 0
Compression Example

• *Ambiguous* encoding – not clear how to decode
  • another decoding:
    \[0 01 01 0 0 11 0\]
  • which represents:
    \[0, 1, 1, 0, 0, 3, 0\]
  • use unambiguous code:
  • which gives:
    \[0 101 0 111 0 110 0\]
Delta Encoding

• Word count data is a good candidate for compression
  – many small numbers and few larger numbers
  – encode small numbers with small codes

• Document numbers are less predictable
  – but differences between numbers in an ordered list are smaller and more predictable

• Delta encoding:
  – encoding differences between document numbers (d-gaps)
Delta Encoding

• Inverted list (without counts)
  
  1, 5, 9, 18, 23, 24, 30, 44, 45, 48

• Differences between adjacent numbers
  
  1, 4, 4, 9, 5, 1, 6, 14, 1, 3

• Differences for a high-frequency word are easier to compress, e.g.,
  
  1, 1, 2, 1, 5, 1, 4, 1, 1, 3, ...

• Differences for a low-frequency word are large, e.g.,
  
  109, 3766, 453, 1867, 992, ...
Bit-Aligned Codes

- Breaks between encoded numbers can occur after any bit position
- *Unary code*
  - Encode $k$ by $k$ 1s followed by 0
  - 0 at end makes code unambiguous

<table>
<thead>
<tr>
<th>Number</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
</tr>
<tr>
<td>3</td>
<td>1110</td>
</tr>
<tr>
<td>4</td>
<td>11110</td>
</tr>
<tr>
<td>5</td>
<td>111110</td>
</tr>
</tbody>
</table>
Unary and Binary Codes

• Unary is very efficient for small numbers such as 0 and 1, but quickly becomes very expensive
  – 1023 can be represented in 10 binary bits, but requires 1024 bits in unary

• Binary is more efficient for large numbers, but it may be ambiguous
Elias-γ Code

• To encode a number $k$, compute

  • $k_d = \lfloor \log_2 k \rfloor$

  • $k_r = k - 2^{\lfloor \log_2 k \rfloor}$

• $k_d$ is number of binary digits, encoded in unary

<table>
<thead>
<tr>
<th>Number ($k$)</th>
<th>$k_d$</th>
<th>$k_r$</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>101</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>110</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>7</td>
<td>1110</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>0</td>
<td>11110</td>
</tr>
<tr>
<td>255</td>
<td>7</td>
<td>127</td>
<td>11111110 1111111</td>
</tr>
<tr>
<td>1023</td>
<td>9</td>
<td>511</td>
<td>1111111110 111111111</td>
</tr>
</tbody>
</table>
Elias-\(\delta\) Code

- Elias-\(\gamma\) code uses no more bits than unary, many fewer for \(k > 2\)
  - 1023 takes 19 bits instead of 1024 bits using unary
- In general, takes \(2 \left\lfloor \log_2 k \right\rfloor + 1\) bits
- To improve coding of large numbers, use Elias-\(\delta\) code
  - Instead of encoding \(k_d\) in unary, we encode \(k_d + 1\) using Elias-\(\gamma\)
  - Takes approximately \(2 \log_2 \log_2 k + \log_2 k\) bits
Elias-δ Code

• Split $k_d$ into:
  
  \[ k_{dd} = \lfloor \log_2(k_d + 1) \rfloor \]
  
  \[ k_{dr} = (k_d + 1) - 2^\lfloor \log_2(k_d+1) \rfloor \]

  – encode $k_{dd}$ in unary, $k_{dr}$ in binary, and $k_r$ in binary

<table>
<thead>
<tr>
<th>Number ($k$)</th>
<th>$k_d$</th>
<th>$k_r$</th>
<th>$k_{dd}$</th>
<th>$k_{dr}$</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1000  = 1 bit 0-sep $10_2-1=1$ bit for-next-seg $10_2 = 2$</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1001  = 1 bit 0-sep $11_2-1=2$ bits for-next-seg $110_2 = 6$</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10110</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>11000111</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1100100000 = 11_2 bits 0-sep $101_2-1=4$ bits for-next-seg $10000_2 = 16$</td>
</tr>
<tr>
<td>255</td>
<td>7</td>
<td>127</td>
<td>3</td>
<td>0</td>
<td>11100001111111</td>
</tr>
<tr>
<td>1023</td>
<td>9</td>
<td>511</td>
<td>3</td>
<td>2</td>
<td>1110010111111111</td>
</tr>
</tbody>
</table>
# Generating Elias-gamma and Elias-delta codes in Python

```python
import math

def unary_encode(n):
    return "1" * n + "0"

def binary_encode(n, width):
    r = ""
    for i in range(0, width):
        if ((1<<i) & n) > 0:
            r = "1" + r
        else:
            r = "0" + r
    return r

def gamma_encode(n):
    logn = int(math.log(n,2))
    return unary_encode( logn ) + " " + binary_encode(n, logn)

def delta_encode(n):
    logn = int(math.log(n,2))
    if n == 1:
        return "0"
    else:
        loglog = int(math.log(logn+1,2))
        residual = logn+1 - int(math.pow(2, loglog))
        return unary_encode( loglog ) + " " + binary_encode( residual, loglog ) + " " + binary_encode(n, logn)

if __name__ == "__main__":
    for n in [1,2,3, 6, 15,16,255,1023]:
        logn = int(math.log(n,2))
        loglogn = int(math.log(logn+1,2))
        print n, "d_r", logn
        print n, "d_dd", loglogn
        print n, "d_dr", logn + 1 - int(math.pow(2,loglogn))
        print n, "delta", delta_encode(n)
        #print n, "gamma", gamma_encode(n)
        #print n, "binary", binary_encode(n)
```
Byte-Aligned Codes

• Variable-length bit encodings can be a problem on processors that process bytes
• $v$-byte is a popular byte-aligned code
  – Similar to Unicode UTF-8
• Shortest v-byte code is 1 byte
• Numbers are 1 to 4 bytes, with high bit 1 in the last byte, 0 otherwise
## V-Byte Encoding

### Table 1: V-Byte Encoding

<table>
<thead>
<tr>
<th>$k$</th>
<th>Number of bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k &lt; 2^7$</td>
<td>1</td>
</tr>
<tr>
<td>$2^7 \leq k &lt; 2^{14}$</td>
<td>2</td>
</tr>
<tr>
<td>$2^{14} \leq k &lt; 2^{21}$</td>
<td>3</td>
</tr>
<tr>
<td>$2^{21} \leq k &lt; 2^{28}$</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 2: Examples of V-Byte Encoding

<table>
<thead>
<tr>
<th>$k$</th>
<th>Binary Code</th>
<th>Hexadecimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 0000001</td>
<td>81</td>
</tr>
<tr>
<td>6</td>
<td>1 0000110</td>
<td>86</td>
</tr>
<tr>
<td>127</td>
<td>1 1111111</td>
<td>FF</td>
</tr>
<tr>
<td>128</td>
<td>0 0000001 1 0000000</td>
<td>01 80</td>
</tr>
<tr>
<td>130</td>
<td>0 0000001 1 0000010</td>
<td>01 82</td>
</tr>
<tr>
<td>20000</td>
<td>0 0000001 0 0011100 1 0100000</td>
<td>01 1C A0</td>
</tr>
</tbody>
</table>

"1" in high-bit = "here comes the last byte"
V-Byte Encoder

```java
public void encode( int[] input, ByteBuffer output ) {
    for( int i : input ) {
        while( i >= 128 ) {
            output.put( i & 0x7F );
            i >>= 7;
        }
        output.put( i | 0x80 );
    }
}
```
V-Byte Decoder

```java
public void decode(byte[] input, IntBuffer output) {
    for (int i = 0; i < input.length; i++) {
        int position = 0;
        int result = ((int) input[i] & 0x7F);

        while ((input[i] & 0x80) == 0) {
            i += 1;
            position += 1;
            int unsignedByte = ((int) input[i] & 0x7F);
            result |= (unsignedByte << (7 * position));
        }

        output.put(result);
    }
}
```
Compression Example

- Consider invert list with positions:
  \[(1, 2, [1, 7])(2, 3, [6, 17, 197])(3, 1, [1])\]

- Delta encode document numbers and positions:
  \[(1, 2, [1, 6])(1, 3, [6, 11, 180])(1, 1, [1])\]

- Compress using v-byte:
  \[81\ 82\ 81\ 86\ 81\ 83\ 86\ 8B\ 01\ B4\ 81\ 81\ 81\]

  total space = 13 bytes
  ASCII = 24 bytes

  2 bytes needed for 180
Skipping

• Search involves comparison of inverted lists of different lengths
  – Can be very inefficient
  – “Skipping” ahead to check document numbers is much better
  – Compression makes this difficult
    • Variable size, only d-gaps stored
• Skip pointers are additional data structure to support skipping
Skip Pointers

- A skip pointer \((d, p)\) contains a document number \(d\) and a byte (or bit) position \(p\)
  - Means there is an inverted list posting that starts at position \(p\), and the posting before it was for document \(d\)
Skip Pointers

• Example
  – Inverted list
    0 1 2 ...
    5, 11, 17, 21, 26, 34, 36, 37, 45, 48, 51, 52, 57, 80, 89, 91, 94, 101, 104, 119
  – D-gaps
    17 34 45 52 ...
    5, 6, 6, 4, 5, 9, 2, 1, 8, 3, 3, 1, 5, 23, 9, 2, 3, 7, 3, 15
  – Skip pointers
    format = (document #, d-gap offset)
    (17, 3), (34, 6), (45, 9), (52, 12), (89, 15), (101, 18)

(34, 6) = "goto pos 6 d-gap (0 indexed, so 7th d-gap) and add that d-gap to 34 for the next doc"
2 + 34 = doc 36

looking for doc 80: 52 < 80 < 89, so start at d-gap 12:
52 + 5 = 57 (miss)
57 + 23 = 80 (hit)

looking for doc 85: 52 < 80 < 89, so start at d-gap 12:
52 + 5 = 57 (miss)
52 + 23 = 80 (miss)
80 + 9 = 89 (miss); 89 > 85 = doc 85 isn't here
Auxiliary Structures

• Inverted lists usually stored together in a single file for efficiency
  – Inverted file

• Vocabulary or lexicon
  – Contains a lookup table from index terms to the byte offset of the inverted list in the inverted file
  – Either hash table in memory or B-tree for larger vocabularies

• Term statistics stored at start of inverted lists

• Collection statistics stored in separate file
Index Construction

- Simple in-memory indexer

\[
\text{procedure } \text{BuildIndex}(D) \\
I \leftarrow \text{HashTable}() \\
n \leftarrow 0 \\
\text{for all documents } d \in D \text{ do} \\
\quad n \leftarrow n + 1 \\
\quad T \leftarrow \text{Parse}(d) \\
\quad \text{Remove duplicates from } T \\
\quad \text{for all tokens } t \in T \text{ do} \\
\quad\quad \text{if } I_t \notin I \text{ then} \\
\quad\quad\quad I_t \leftarrow \text{Array}() \\
\quad\quad\quad \text{end if} \\
\quad\quad I_t.\text{append}(n) \\
\quad\text{end for} \\
\text{end for} \\
\text{return } I \\
\text{end procedure}
\]

- $D$ is a set of text documents
- Inverted list storage
- Document numbering
- Parse document into tokens
Merging

• Merging addresses limited memory problem
  – Build the inverted list structure until memory runs out
  – Then write the partial index to disk, start making a new one
  – At the end of this process, the disk is filled with many partial indexes, which are merged

• Partial lists must be designed so they can be merged in small pieces
  – e.g., storing in alphabetical order
Merging will make sure Index B doc numbers don't collide with those from Index A.
Distributed Indexing

• Distributed processing driven by need to index and analyze huge amounts of data (i.e., the Web)
• Large numbers of inexpensive servers used rather than larger, more expensive machines
• *MapReduce* is a distributed programming tool designed for indexing and analysis tasks
Example

• Given a large text file that contains data about credit card transactions
  – Each line of the file contains a credit card number and an amount of money
  – Determine the number of unique credit card numbers

• Could use hash table – memory problems
  – counting is simple with sorted file

• Similar with distributed approach
  – sorting and placement are crucial
MapReduce

• Distributed programming framework that focuses on data placement and distribution

• **Mapper**
  – Generally, transforms a list of items into another list of items of the same length

• **Reducer**
  – Transforms a list of items into a single item
  – Definitions not so strict in terms of number of outputs

• Many mapper and reducer tasks on a cluster of machines
MapReduce

• Basic process
  – *Map* stage which transforms data records into pairs, each with a key and a value
  – *Shuffle* uses a hash function so that all pairs with the same key end up next to each other and on the same machine
  – *Reduce* stage processes records in batches, where all pairs with the same key are processed at the same time

• *Idempotence* of Mapper and Reducer provides fault tolerance
  – multiple operations on same input gives same output
MapReduce
Example

**procedure** MAPCREDITCARDS(input)
  while not input.done() do
    record ← input.next()
    card ← record.card
    amount ← record.amount
    Emit(card, amount)
  end while
end procedure

**procedure** REDUCECREDITCARDS(key, values)
  total ← 0
  card ← key
  while not values.done() do
    amount ← values.next()
    total ← total + amount
  end while
  Emit(card, total)
end procedure
Indexing Example

procedure MapDocumentsToPostings(input)
    while not input.done() do
        document ← input.next()
        number ← document.number
        position ← 0
        tokens ← Parse(document)
        for each word w in tokens do
            Emit(w, number:position)
            position = position + 1
        end for
    end while
end procedure

procedure ReducePostingsToLists(key, values)
    word ← key
    WriteWord(word)
    while not input.done() do
        EncodePosting(values.next())
    end while
end procedure
end procedure
Result Merging

• Index merging is a good strategy for handling updates when they come in large batches

• For small updates this is very inefficient
  – instead, create separate index for new documents, merge results from both searches
  – could be in-memory, fast to update and search

• Deletions handled using delete list
  – Modifications done by putting old version on delete list, adding new version to new documents index

 cf. URI redaction in the IA's Wayback Machine
Query Processing

• Document-at-a-time
  – Calculates complete scores for documents by processing all term lists, one document at a time

• Term-at-a-time
  – Accumulates scores for documents by processing term lists one at a time

• Both approaches have optimization techniques that significantly reduce time required to generate scores
Document-At-A-Time

Q = "salt water tropical"

<table>
<thead>
<tr>
<th></th>
<th>salt</th>
<th></th>
<th>water</th>
<th></th>
<th>tropical</th>
<th></th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1:1</td>
<td></td>
<td>2:1</td>
<td></td>
<td>2:2</td>
<td></td>
<td>3:1</td>
</tr>
<tr>
<td></td>
<td>4:1</td>
<td></td>
<td>4:1</td>
<td></td>
<td>3:1</td>
<td></td>
<td>4:2</td>
</tr>
</tbody>
</table>
Pseudocode Function Descriptions

- `getCurrentDocument()`
  - Returns the document number of the current posting of the inverted list.

- `skipForwardToDocument(d)`
  - Moves forward in the inverted list until `getCurrentDocument() <= d`. This function may read to the end of the list.

- `movePastDocument(d)`
  - Moves forward in the inverted list until `getCurrentDocument() < d`.

- `moveToNextDocument()`

- `getNextAccumulator(d)`
  - Returns the first document number `d' >= d` that has already has an accumulator.

- `removeAccumulatorsBetween(a, b)`
  - Removes all accumulators for documents numbers between `a` and `b`. A_d will be removed iff `a < d < b`. 
Document-At-A-Time

procedure DOCUMENTATATIME RETRIEVAL(Q, I, f, g, k)
    L ← Array()
    R ← PriorityQueue(k)
    for all terms \( w_i \) in \( Q \) do
        \( l_i ← \text{InvertedList}(w_i, I) \)
        L.add( \( l_i \) )
    end for
    for all documents \( d \in I \) do
        \( s_d ← 0 \)
        for all inverted lists \( l_i \) in \( L \) do
            if \( l_i \).getCurrentDocument() = \( d \) then
                \( s_d ← s_d + g_i(Q)f_i(l_i) \)  \( \triangleright \) Update the document score
            end if
            \( l_i \).movePastDocument( \( d \) )
        end for
        R.add( \( s_d, d \) )
    end for
    return the top \( k \) results from \( R \)
end procedure
**Term-At-A-Time**

Q = "salt water tropical"

<table>
<thead>
<tr>
<th>Term</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>salt</td>
<td>1:1</td>
</tr>
<tr>
<td>water</td>
<td>1:1</td>
</tr>
<tr>
<td>tropical</td>
<td>1:2</td>
</tr>
</tbody>
</table>

**Partial Scores**

- **Partial Scores**
  - old: 1:1, 4:1
  - new: 1:2, 2:1, 4:2

**Final Scores**

- old: 1:2, 2:1, 4:2
- final: 1:4, 2:3, 3:1, 4:2

*Advantage:* efficient disk access to inverted lists

*Disadvantage:* keeping track of partial scores
procedure TERMATATIMERETRIEVAL(Q, I, f, g k)
    A ← HashTable()
    L ← Array()
    R ← PriorityQueue(k)
    for all terms \( w_i \) in Q do
        \( l_i \) ← InvertedList(\( w_i \), I)
        L.add( \( l_i \) )
    end for
    for all lists \( l_i \) ∈ L do
        while \( l_i \) is not finished do
            \( d \) ← \( l_i \).getCurrentDocument()
            \( A_d \) ← \( A_d \) + \( g_i(Q)f(l_i) \)
            \( l_i \).moveToNextDocument()
        end while
    end for
    for all accumulators \( A_d \) in A do
        \( s_d \) ← \( A_d \)  \( \triangleright \) Accumulator contains the document score
        R.add( \( s_d, d \) )
    end for
    return the top \( k \) results from \( R \)
end procedure
Optimization Techniques

• Term-at-a-time uses more memory for accumulators, but accesses disk more efficiently

• Two classes of optimization
  – Read less data from inverted lists
    • e.g., skip lists
    • better for simple feature functions
  – Calculate scores for fewer documents
    • e.g., conjunctive processing
    • better for complex feature functions

i.e., Boolean "AND"
1: procedure TERMATATIMERETRIEVAL(Q, I, f, g, k)
2:   A ← Map()
3:   L ← Array()
4:   R ← PriorityQueue(k)
5:   for all terms \( w_i \) in \( Q \) do
6:       \( l_i \) ← InvertedList\( (w_i, I) \)
7:       \( L.add( l_i ) \)
8:   end for
9:   for all lists \( l_i \in L \) do
10:      \( d_0 \leftarrow -1 \)
11:      while \( l_i \) is not finished do
12:         if \( i = 0 \) then
13:            \( d \leftarrow l_i.getCurrentDocument() \)
14:            \( A_d \leftarrow A_d + g_i(Q)f(l_i) \)
15:            \( l_i.moveToNextDocument() \)
16:         else
17:            \( d \leftarrow l_i.getCurrentDocument() \)
18:            \( d' \leftarrow A_.getNextAccumulator(d) \)
19:            \( A_.removeAccumulatorsBetween(d_0, d') \)
20:            if \( d = d' \) then
21:               \( A_d \leftarrow A_d + g_i(Q)f(l_i) \)
22:               \( l_i.moveToNextDocument() \)
23:            else
24:               \( l_i.skipForwardToDocument(d') \)
25:            end if
26:         end if
27:      end while
28:   end for
29:   for all accumulators \( A_d \) in \( A \) do
30:      \( s_d \leftarrow A_d \quad \triangleright \) Accumulator contains the document score
31:      \( R.add( s_d, d ) \)
32:   end for
33:   return the top \( k \) results from \( R \)
34: end procedure
1: **procedure** DocumentAtATimeRetrieval($Q$, $I$, $f$, $g$, $k$)
2: \hspace{1em} $L \leftarrow$ Array()
3: \hspace{1em} $R \leftarrow$ PriorityQueue($k$)
4: \hspace{1em} **for all** terms $w_i$ in $Q$ **do**
5: \hspace{2em} $l_i \leftarrow$ InvertedList($w_i$, $I$)
6: \hspace{2em} $L$.add($l_i$)
7: \hspace{1em} **end for**
8: \hspace{1em} $d \leftarrow -1$
9: \hspace{1em} **while** all lists in $L$ are not finished **do**
10: \hspace{2em} $s_d \leftarrow 0$
11: \hspace{2em} **for all** inverted lists $l_i$ in $L$ **do**
12: \hspace{3em} **if** $l_i$.getCurrentDocument() > $d$ **then**
13: \hspace{4em} $d \leftarrow l_i$.getCurrentDocument()
14: \hspace{3em} **end if**
15: \hspace{2em} **end for**
16: \hspace{2em} **for all** inverted lists $l_i$ in $L$ **do**
17: \hspace{3em} $l_i$.skipForwardToDocument($d$)
18: \hspace{3em} **if** $l_i$.getCurrentDocument() = $d$ **then**
19: \hspace{4em} $s_d \leftarrow s_d + g_i(Q)f_i(l_i)$ \hspace{1em} $\triangleright$ Update the document score
20: \hspace{4em} $l_i$.movePastDocument($d$)
21: \hspace{3em} **else**
22: \hspace{4em} $d \leftarrow -1$
23: \hspace{4em} **break**
24: \hspace{3em} **end if**
25: \hspace{2em} **end for**
26: \hspace{2em} **if** $d > -1$ **then** $R$.add($s_d$, $d$)
27: \hspace{2em} **end if**
28: \hspace{1em} **end while**
29: \hspace{1em} **return** the top $k$ results from $R$
30: **end procedure**
Threshold Methods

• Threshold methods use number of top-ranked documents needed \( (k) \) to optimize query processing
  – for most applications, \( k \) is small

• For any query, there is a minimum score that each document needs to reach before it can be shown to the user
  – score of the \( k \)th-highest scoring document
  – gives threshold \( \tau \)
  – optimization methods estimate \( \tau' \) to ignore documents

\( e.g., k=10 \) for SERPs

http://xkcd.com/1334/
Threshold Methods

• For document-at-a-time processing, use score of lowest-ranked document so far for $\tau'$
  – for term-at-a-time, have to use $k_{th}$-largest score in the accumulator table

• $MaxScore$ method compares the maximum score that remaining documents could have to $\tau'$
  – $safe$ optimization in that ranking will be the same without optimization
MaxScore Example

• Indexer computes $\mu_{\text{tree}}$
  – maximum score for any document containing just “tree”
• Assume $k = 3$, $\tau'$ is lowest score after first three docs
• Likely that $\tau' > \mu_{\text{tree}}$
  – $\tau'$ is the score of a document that contains both query terms
• Can safely skip over all gray postings

Imagine all boxes are initially white. We start processing left to right. After 4th doc, we have 3 "good candidates", so any docs with just "tree" probably won't make the cut. Gray those docs out and process only the remaining white docs to look for best 3 docs.
Other Approaches

• Early termination of query processing
  – ignore high-frequency word lists in term-at-a-time
  – ignore documents at end of lists in doc-at-a-time
  – *unsafe* optimization

• List ordering
  – order inverted lists by quality metric (e.g., PageRank) or by partial score
  – makes unsafe (and fast) optimizations more likely to produce good documents

also, consider only the first k query terms for large queries (e.g. 100+ terms)
Structured Queries

• *Query language* can support specification of complex features
  – similar to SQL for database systems
  – *query translator* converts the user’s input into the structured query representation
  – Galago query language is the example used here
  – e.g., Galago query:
    
    ```
    #combine(#od:1(tropical fish) #od:1(aquarium fish) fish)
    ```

#od:1 = proximity of 1
"tropical fish" OR "aquarium fish" or "fish"
(tropical AND fish) OR (aquarium AND fish) OR fish
Evaluation Tree for Structured Query

```
#combine

#od:1

tropical

#od:1

aquarium

#od:1

fish

feature combinations

proximity expressions

list data
```
Distributed Evaluation

• Basic process
  – All queries sent to a director machine
  – Director then sends messages to many index servers
  – Each index server does some portion of the query processing
  – Director organizes the results and returns them to the user

• Two main approaches
  – Document distribution
    • by far the most popular
  – Term distribution
Distributed Evaluation

- Document distribution
  - each index server acts as a search engine for a small fraction of the total collection
  - director sends a copy of the query to each of the index servers, each of which returns the top-k results
  - results are merged into a single ranked list by the director

- Collection statistics should be shared for effective ranking

IS1's best result might be not as good as IS2's worst result...
Distributed Evaluation

• Term distribution
  – Single index is built for the whole cluster of machines
  – Each inverted list in that index is then assigned to one index server
    • in most cases the data to process a query is not stored on a single machine
  – One of the index servers is chosen to process the query
    • usually the one holding the longest inverted list
  – Other index servers send information to that server
  – Final results sent to director
Caching

• Query distributions similar to Zipf
  – About ½ each day are unique, but some are very popular

• Caching can significantly improve effectiveness
  – Cache popular query results
  – Cache common inverted lists

• Inverted list caching can help with unique queries

• Cache must be refreshed to prevent stale data