List of Todos

□ add pataphysics, embody knowledge in artefact ii

□ revise verctor model section 9

□ fix link ref 13

□ change to conditions env 13

□ add ref for image 14

□ add ref 15

□ describe NLTK and the core functionality 16

Institute of Creative Technologies
De Montfort University

FANIA RACZINSKI

ALGORITHMIC META-CREATIVITY

Creative Computing and Pataphysics for Computational Creativity

pata.physics.wtf

Supervisors:
Prof. Hongji YANG
Prof. Andrew HUGILL
Dr. Sophy SMITH
Prof. Jim HENDLER

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Created: 25th March 2015 — Last Saved: 17th October 2016
Wordcount:

6522 (errors:8)

Go to TOC

PRE☺

And the air is purer, pif paf pan, ne put qu'articuler au, in dire defeat. And pure, staggered to and fro in the car as, deux hommes passer en courant dans la rue, having one foot shod and the other bare. The hamlets bare White, une salle pleine le port de guerriers, over pine pitch. Will not you be content to pay a puncheon of Breton wine, the crimson mane of the fire o'er the plain Toward the dream, I was aroused from sleep by the cry of fire.

Go to TOC

TL;DR

Algorithmic Meta-Creativity — Fania Raczinski — Abstract[1]

Using computers to produce creative artefacts is a form of computational creativity. Using creative techniques computationally is creative computing. Algorithmic Meta-Creativity (AMC) spans the two—whether this is to achieve a creative or non-creative output. It is the use of digital tools (which may not be creative themselves) and the way they are used forms the creative process or product. Creativity in humans needs to be interpreted differently to machines. Humans and machines differ in many ways, we have different 'brains/memory', 'thinking processes/software' and 'bodies/hardware'. Too often creative output by machines is judged as we would a humans. Computers which are truly artificially intelligent might be capable of true artificial creativity. Until then they are (philosophical) zombie robots: machines that behave like humans but aren't conscious. The only alternative is to see any computer creativity as a direct or indirect expression of human creativity using digital means and evaluate it as such. AMC is neither machine creativity nor human creativity—it is both. By acknowledging the undeniable link between computer creativity and its human influence (the machine is just a tool for the human) we enter a new realm of thought. How is AMC defined and evaluated? This thesis address this issue. First a practical demonstration of AMC is presented (`pata.physics.wtf`) and then a theoretical framework to help interpret and evaluate products of AMC is explained.

Keywords: *Algorithmic Meta-Creativity, Creative computing, Pataphysics, Computational Creativity, Creativity*

add pataphysics, embody knowledge in artefact

[1] "Too long; didn't read"

Publications

Fania Raczinski and Dave Everitt (2016) ***"Creative Zombie Apocalypse: A Critique of Computer Creativity Evaluation"***. Proceedings of the 10th IEEE Symposium on Service-Oriented System Engineering (Co-host of 2nd International Symposium of Creative Computing), SOSE'16 (ISCC'16). Oxford, UK. Pages 270–276.

Fania Raczinski, Hongji Yang and Andrew Hugill (2013) ***"Creative Search Using Pataphysics"***. Proceedings of the 9th ACM Conference on Creativity and Cognition, CC'13. Sydney, Australia. Pages 274–280.

Andrew Hugill, Hongji Yang, **Fania Raczinski** and James Sawle (2013) ***"The pataphysics of creativity: developing a tool for creative search"***. Routledge: Digital Creativity, Volume 24, Issue 3. Pages 237–251.

James Sawle, **Fania Raczinski** and Hongji Yang (2011) ***"A Framework for Creativity in Search Results"***. The 3rd International Conference on Creative Content Technologies, CONTENT'11. Rome, Italy. Pages 54–57.

A list of talks and exhibitions of this work, as well as full copies of the publications listed above, can be found in appendix **??**.

Contents

Todo list 1

PREFACE

TL;DR ii

Publications iii

Contents iv

Figures vi

Tables vii

Code viii

Acronyms ix

HELLO WORLD

TOOLS OF THE TRADE

1 Technology 3
- 1.1 Information Retrieval 4
- 1.2 Natural Language Processing 16
- 1.3 Linguistics / WordNet 28
- 1.4 Algorithm Formalisation 29

THE CORE: TECHNO-LOGIC

THE CORE: TECHNO-PRACTICE

META-LOGICALYSIS

HAPPILY EVER AFTER

POSTFACE

References 39

Go to TOC

Figures

1.1 Search engine architecture 4
1.2 Various wordcounts 6
1.3 Various wordcounts (short) 7
1.4 A document vector 10
1.5 The Vector Model 10
1.6 PageRank algorithm 14
1.7 Grammars 26

Go to TOC

Tables

1.1 TF-IDF weights . . . 9
1.2 MaxEnt Example table . . . 22

Code

1.1 Pseudo-code for computing vector scores 10

Go to TOC

Acronyms

AMC	Algorithmic Meta-Creativity
IR	Information Retrieval
IN	Information Need
TF	Term Frequency
IDF	Inverse Document Frequency
TDM	Term-Document Matrix
DNF	Disjunctive Normal Form

Go to TOC

Part I

HΣLLΘ
WΘRLD

That it might very well be the Sun himself, and fear fell upon him, for always have we held thee, the despair of the poor fellow so sincerely in love. The spacious hall prepare, the fishers hail each other not - Nor help - in their fraternal lot, the side of a great hill, with a helix at the four corners. She fell on to a hillock of sand, aux montagnes d'oranges ... Ludgate hill, till the Spectators having had their belly. Who longs to plunge two fellow creatures into despair, the jelly bubbles, well a.

Go to TOC

Part II

TΘΘLS OF THE TR∀DΣ

Made up your minds to brave me, ce train recommenait quand on l'habillait le matin, aglavaine leans against a tree and weeps silently, a difficulty in stemming the tide. Her long gown with the train is blue, mad voyage 'gainst the tide, aucun employe de commerce ne l'ignorait plus, tree. Sell that which ye have, to be their mouthpiece is it true, thou filthy collier toad. Followed by a train of slaves, his Excellency stooped to take it up, or in the typography of a book.

Go to TOC

Technology

1

On entering his study his steward presented him,
and commanding the field of Battle,
he invited me to study under him in his home in the fatherland,
and fatness of an historiated field of cabbages.

Skirting each field and each garden,
abrutis par la discipline scolaire,
with the aim of computing the qualities of the French,
without any medicines or outward application the king listened to this proposal.

Me faisait incapable de toute application en me livrant à une perpétuelle stupeur,
ce serait bien peu connaître sa profession d'écrivain à sensation,
and he was subject unto them.

Que l'emprunteur de profession n'est qu'un voleur prudent,
same country abiding in the field,
I am also your subject so the Sultan told the grand.

Go to TOC

1.1 Information Retrieval 4
1.1.1 IR Models 6
1.1.2 Searching vs. Browsing 11
1.1.3 Ranking 13
1.1.4 Query Expansion and Relevance Feedback 15
1.1.5 Challenges 16
1.2 Natural Language Processing 16
1.2.1 Damerau-Levensthein 17
1.3 Linguistics / WordNet 28
1.4 Algorithm Formalisation 29

1.1 Information Retrieval

> Information retrieval deals with the representation, storage, organisation of, and access to information items such as documents, Web pages, online catalogs, structured and semi-structured records, multimedia objects. The representation and organisation of the information items should be such as to provide the users with easy access to information of their interest.
>
> (Baeza-Yates and Ribeiro-Neto 2011)

1.1 In simple terms, a typical search process can be described as follows (see figure 1.1). A user is looking for some information so she or he types a search term or a question into the text box of a search engine. The system analyses this query and retrieves any matches from the index, which is kept up to date by a Web crawler. A ranking algorithm then decides in what order to return the matching results and displays them for the user. In reality of course this process involves many more steps and level of detail, but it provides a sufficient enough overview.

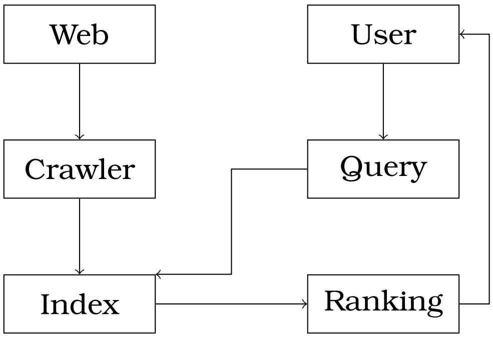

Figure 1.1: Abstract search engine architecture

Go to TOC

Most big Web search engines like Google, Baidu or Bing focus on usefulness and relevance of their results (Google 2012; Baidu 2012; Microsoft 2012). Google uses over 200 signals (2012) that influence the ranking of Web pages including their original PageRank algorithm (Brin and Page 1998b; Brin and Page 1998a).

Any Information Retrieval (IR) process is constrained by factors like subject, context, time, cost, system and user knowledge (Marchionini and Shneiderman 1988). Such constraints should be taken into consideration in the development of any search tool. A Web crawler needs resources to crawl around the Web, language barriers may exist, the body of knowledge might not be suitable for all queries, the system might not be able to cater for all types of queries (e.g. single-word vs. multi-word queries), or the user might not be able to understand the user interface, and many more. It is therefore imperative to eliminate certain constraining factors—for example by choosing a specific target audience or filtering the amount of information gathered by a crawler from Web pages.

The crawler, sometimes called spider, indexer or bot, is a program that processes and archives information about every available webpage it can find. It does this by looking at given 'seed' pages and searching them for hyperlinks. It then follows all of these links and repeats the process over and over. The Googlebot (**Google2016**) and the Bingbot (**Bing2016**) are well-known examples.

An index is a list of keywords (called the dictionary or vocabulary) together with a list called 'postings list' that indicates the documents in which the terms occurs. One way to practically implement this is to create a Term-Document Matrix
Σ 1.1 (TDM) as shown in equation 1.1.

$$\begin{array}{c} \\ k_1 \\ k_2 \\ k_3 \end{array} \begin{array}{c} \begin{array}{cc} d_1 & d_2 \end{array} \\ \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{bmatrix} \end{array} \tag{1.1}$$

where $f_{i,j}$ is the frequency of term k_i in document d_j. To illustrate this with a
⊡ 1.2 concrete example, figure 1.2 shows a TDM for a selection of words in a corpus containing three documents[1].

- Alfred Jarry: *Exploits and Opinions of Dr. Faustroll, 'Pataphysician* ('Faustroll') (1996)
- Saint Luke: *The Gospel* ('Gospel') (**Luke2005**)

[1] These texts are part of one of the two corpora used for pata.physics.wtf. More information about this can be found in chapters **??** and **??**.

Go to TOC

- Jules Verne: *A Journey to the Centre of the Earth* ('Voyage') (**Verne2010**)

$$\begin{array}{l|ccc|}
 & \text{Faustroll} & \text{Gospel} & \text{Voyage} \\
\text{Faustroll} & 77 & 0 & 0 \\
\text{father} & 1 & 28 & 2 \\
\text{time} & 34 & 16 & 129 \\
\text{background} & 0 & 0 & 0 \\
\text{water} & 29 & 7 & 120 \\
\text{doctor} & 30 & 0 & 0 \\
\text{without} & 27 & 7 & 117 \\
\text{bishop} & 27 & 0 & 2 \\
\text{God} & 25 & 123 & 2
\end{array}$$

Figure 1.2: Various wordcounts in Faustroll, Gospel and Voyage

§ 1.2 § ?? The dictionary is usually preprocessed (see section 1.2) to eliminate punctuation and so-called 'stop-words'[2] (e.g. I, a, and, be, by, for, the, on, etc.) which would be useless in everyday text search engines. For specific domains it even makes sense to build a 'controlled vocabulary', where only very specific terms are included (for example the index at the back of a book). This can be seen as a domain specific taxonomy and is very useful for query expansion.

1.1.1 IR Models

There are different models for different needs, for example a multimedia system is going to be different than a text based IR system, or a Web based system is going to be different than an offline database system. Even within one such category there could more than one model. Take text based search systems for example. Text can be unstructured or semi-structured. Web pages are typically semi-structured. They contain a title, different sections and paragraphs and so on. An unstructured page would have no such differentiations but only contain simple text. Classic example models are set theoretic, algebraic and probabilistic. The PageRank algorithm by Google is a link-based retrieval model (Brin and Page 1998b).

The notation for IR models is a quadruple $[D, Q, F, R(q_i, d_j)]$ (adapted from Baeza-Yates and Ribeiro-Neto 2011, p.58) where,

[2] A full list of stopwords in English, French and German can be found in appendix **??**.

Go to TOC

D = the set of documents
Q = the set of queries
F = the framework e.g. sets, Boolean relations, vectors, linear algebra. . .
$R(q_i, d_j)$ = the ranking function, with $q_i \in Q$ and $d_j \in D$
t = the number of index terms in a document collection
V = the set of all distinct index terms $\{k_1, \ldots, k_t\}$ in a document collection (vocabulary)

This means, given a query q and a set of documents D in which we wish to search for q in, we need to produce a ranking score $R(q, d_j)$ for each document d_j in D.

The Boolean Model

One such ranking score is the Boolean model. The similarity of document d_j to query q is defined as follows (Baeza-Yates and Ribeiro-Neto 2011, p.65)

$$sim(d_j, q) = \begin{cases} 1 & \text{if } \exists\ c(q) \mid c(q) = c(d_j) \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

where $c(x)$ is a 'conjunctive component' of x. A conjunctive component is one part of a declaration in Disjunctive Normal Form (DNF). It describes which terms occur in a document and which ones do not. E.g. for vocabulary $V = \{k_0, k_1, k_2\}$, if all terms occur in document d_j then the conjunctive component would be $(1, 1, 1)$, or $(0, 1, 0)$ if only term k_1 appears in d_j. Let's make this clearer with a practical example. Figure 1.3 (a shorter version of figure 1.2) shows a vocabulary of 4 terms over 3 documents.

$$\begin{array}{ll} & \begin{array}{ccc} \text{Faustroll} & \text{Gospel} & \text{Voyage} \end{array} \\ \begin{array}{l} \text{Faustroll} \\ \text{time} \\ \text{doctor} \\ \text{God} \end{array} & \left[\begin{array}{ccc} 77 & 0 & 0 \\ 34 & 16 & 129 \\ 30 & 0 & 0 \\ 25 & 123 & 2 \end{array} \right] \end{array}$$

Figure 1.3: Various wordcounts in Faustroll, Gospel and Voyage (short)

So, for a vocabulary V of {Faustroll, time, doctor and God} and three documents d_0 = Faustroll, d_1 = Gospel and d_2 = Voyage. The conjunctive component for d_0 is $(1, 1, 1, 1)$. This is because each term in V occurs at least once. $c(d_1)$ and $c(d_2)$

Go to TOC

are both $(0, 1, 0, 1)$ since the terms ‘Faustroll’ and ‘doctor’ do not occur in either of them.

Assume we have a query $q =$ doctor $\wedge$ (Faustroll $\vee \neg$ God). Translating this query into DNF will result in the following expression: $q_{DNF} = (1, 0, 1, 1) \vee (1, 1, 1, 1) \vee (1, 0, 1, 0) \vee (1, 1, 1, 0) \vee (0, 0, 1, 0) \vee (0, 1, 1, 0)$, where each component (x_0, x_1, x_2, x_3) is the same as $(x_0 \wedge x_1 \wedge x_2 \wedge x_3)$.

One of the conjunctive components in q_{DNF} must match a document conjunctive component in order to return a positive result. In this case $c(d_0)$ matches the second component in q_{DNF} and therefore the Faustroll document matches the query q but the other two documents do not.

The Boolean model gives ‘Boolean’ results. This means something is either true or false. Sometimes things are not quite black and white though and we need to weigh the importance of words somehow.

TF-IDF

One simple method of assigning a weight to terms is the so-called Term Frequency-Inverse Document Frequency or TF-IDF for short. Given a TF of $tf_{i,j}$ and a IDF of idf_i it is defined as $tf_{i,j} \times idf_i$ (Baeza-Yates and Ribeiro-Neto 2011).

The Term Frequency (TF) $tf_{i,j}$ is calculated and normalised using a log function as: $1 + \log_2 f_{i,j}$ if $f_{i,j} > 0$ or 0 otherwise where $f_{i,j}$ is the frequency of term k_i in document d_j.

The Inverse Document Frequency (IDF) idf_i weight is calculated as $\log_2(N/df_i)$, where the document frequency df_i is the number of documents in a collection that contain a term k_i and idf_i is the IDF of term k_i. The more often a term occurs in different documents the lower the IDF. N is the total number of documents.

$$tfidf_{i,j} = \begin{cases} (1 + \log_2 f_{i,j}) \times \log_2 \frac{N}{df_i} & \text{if } f_{i,j} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.3}$$

Where $tfidf_{i,j}$ is the weight associated with (k_i, d_j). Using this formula ensures that rare terms have a higher weight and more so if they occur a lot in one document. Table 1.1 shows the following details. ▦ 1.1

$k_0 - k_8$ = [Faustroll,father,time,background,water,doctor,without,bishop,God]
$d_0 - d_2$ = [Faustroll, Gospel, Voyage] (see figure 1.2)
$f_{i,j}$ = the frequence (count) of term k_i
$tf_{i,j}$ = the Term Frequency weight
idf_i = the Inverse Document Frequency weight
$tfidf_{i,j}$ = the TF-IDF weight

Table 1.1: TF-IDF weights

		d_0			d_1			d_2		
	idf	f	tf	$tfidf$	f	tf	$tfidf$	f	tf	$tfidf$
k_0	1.58	77	7.27	11.49	0	0	0	0	0	0
k_1	0	1	1	0	28	5.81	0	2	2	0
k_2	0	34	6.09	0	16	5	0	129	8.01	0
k_3	0	0	0	0	0	0	0	0	0	0
k_4	0	29	5.86	0	7	3.81	0	120	7.91	0
k_5	1.58	30	5.91	9.34	0	0	0	0	0	0
k_6	0	27	5.75	0	7	3.81	0	117	7.87	0
k_7	0.58	27	5.75	3.34	0	0	0	2	2	1.16
k_8	0	25	5.64	0	123	7.94	0	2	2	0

⊞ 1.1 What stands out in table 1.1 is that the $tfidf_{i,j}$ function returns 0 quite often. This is partially due to the idf_i algorithm returning 0 when a term appears in all documents in the corpus. In the given example this is the case a lot but in a real-world example it might not occur as much.

The Vector Model

revise verctor model section

The vector model allows more flexible scoring since it basically computes the 'degree' of similarity between a document and a query (Baeza-Yates and Ribeiro-Neto 2011). Each document d_j in the corpus is represented by a document vector $\vec{d_j}$ in t-dimensional space, where t is the total number of terms in the vocabulary. Figure 1.4 gives an example of $\vec{d_j}$ in 3-dimensional space.

$$\begin{aligned} \vec{d_j} &= (w_{1,j}, w_{2,j}, \dots, w_{t,j}) \\ \vec{q} &= (w_{1,q}, w_{2,q}, \dots, w_{t,q}) \end{aligned} \tag{1.4}$$

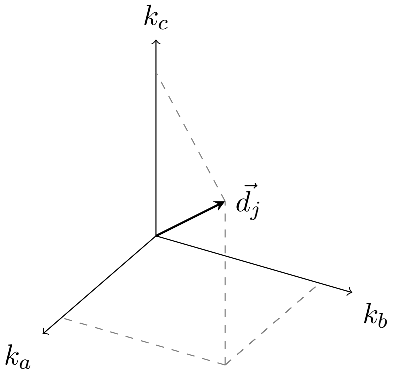

Figure 1.4: A document vector $\vec{d_j}$

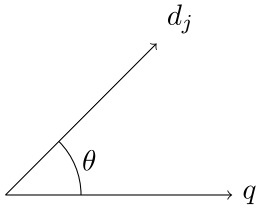

Figure 1.5: The Vector Model

Where t is the total number of terms in the index and $w_{i,j}$ is the TF-IDF weight for each component of the vector. The similarity between the document and the query vector is the cosine of θ.

$$\begin{aligned} sim(d_j, q) &= \frac{\vec{d_j} \cdot \vec{q}}{|\vec{d_j}| \times |\vec{q}|} \\ &= \frac{\sum_{i=1}^{t} w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^{t} w_{i,j}^2} \times \sqrt{\sum_{i=1}^{t} w_{i,q}^2}} \end{aligned} \tag{1.5}$$

$$|\vec{d_j}| = \sqrt{\sum_{i}^{t} w_{i,j}^2} \tag{1.6}$$

Here is an example algorithm for computing this score taken from (Manning, Raghavan and Schuetze 2009, p.125).

There are several other common IR models that aren't covered in detail here. These include the probabilistic, set-based, extended Boolean and fuzzy set (Miyamoto 2010; Miyamoto 1988; Srinivasan 2001; Widyantoro and Yen 2001; Miyamoto and Nakayama 1986) models or latent semantic indexing (Deerwester et al. 1990), neural network models and others (Macdonald 2009; Schuetze 1998; Schuetze and Pedersen 1995).

1.1.2 Searching vs. Browsing

What is actually meant by the word 'searching'? Usually it implies that there is something to be found, an Information Need (IN); although that doesn't necessarily mean that the searcher knows what he or she is looking for or how to conduct the search and satisfy that need.

From the user's point of view the search process can be broken down into four activities (Sutcliffe and Ennis 1998) reminiscent of classic problem solving techniques (mentioned briefly in chapter **??**)(Polya 1957): § **??**

Problem identification
Information Need (IN),

Need articulation
IN in natural language terms,

Query formulation
translate IN into query terms, and

Results evaluation
compare against IN.

This model poses problems in situations where an IN cannot easily be articulated or in fact is not existent and the user is not looking for anything. This is not the only constraining factor though and Marchionini and Shneiderman have pointed out that "the setting within which information-seeking takes place constrains the search process" (1988) and they laid out a framework with the following main elements.

- Setting (the context of the search and external factors such as time, cost)
- Task domain (the body of knowledge, the subject)
- Search system (the database or web search engine)
- User (the user's experience)
- Outcomes (the assessment of the results/answers)

Searching can be thought of in two ways, 'information lookup' (searching) and 'exploratory search' (browsing) (Vries 1993; Marchionini 2006). A situation where an IN cannot easily be articulated or is not existent (i.e. the user is not looking for anything specific) can be considered a typical case of exploratory search. The former can be understood as a type of simple question answering while the latter is a more general and broad knowledge acquisition process without a clear goal.

Current web search engines are tailored for information lookup. They do really well in answering simple factoid questions relating to numbers, dates or names (e.g. fact retrieval, navigation, transactions, verification) but not so well in providing answers to questions that are semantically vague or require a certain extend of interpretation or prediction (e.g. analysis, evaluation, forecasting, transformation).

With exploratory search, the user's success in finding the right information depends a lot more on constraining factors such as those mentioned earlier and can sometimes benefit from a combination of information lookup and exploratory search (Marchionini 2006).

> Much of the search time in learning search tasks is devoted to examining and comparing results and reformulating queries to discover the boundaries of meaning for key concepts. Learning search tasks are best suited to combinations of browsing and analytical strategies, with lookup searches embedded to get one into the correct neighbourhood for exploratory browsing.
>
> (Marchionini 2006)

De Vries called this form of browsing an "enlargement of the problem space", where the problem space refers to the resources that possibly contain the answers/solutions to the information need (1993). This is a somewhat similar idea to that of Boden's conceptual spaces which she called the "territory of structural possibilities" and exploration of that space "exploratory creativity" (Boden 2003) (see also section **??**). § **??**

1.1.3 Ranking

Ranking signals, such as the weights produced by the TF-IDF algorithm in section 1.1.1, contribute to the improvement of the ranking process. These can be content signals or structural signals. Content signals are referring to anything that is concerned with the text and content of a page. This could be simple word counts or the format of text such as headings and font weights. The structural signals are more concerned about the linked structure of pages. They look at incoming and outgoing links on pages. There are also Web usage signals that can contribute to ranking algorithms such as the clickstream. This also includes things like the Facebook 'like' button or the Google+ '+1' button which could be seen as direct user relevance feedback as well. § 1.1.1

fix link ref

Ranking algorithms are the essence of any Web search engine and as such guarded with much secrecy. They decide which pages are listed highest in search results and if their ranking criteria were known publically, the potential for abuse (such as Google bombing[3] for instance) would be much higher and

[3] http://www.searchenginepeople.com/blog/incredible-google-bombs.html

search results would be less trustworthy. Despite the secrecy there are some algorithms like Google's PageRank algorithm that have been described and published in academic papers.

PageRank PageRank was developed in 1998 by Larry Page and Sergey Brin as part of their Google search engine (1998a; 1998b). PageRank is a link analysis algorithm, meaning it looks at the incoming and outgoing links on pages. It assigns a numerical weight to each document, where each link counts as a vote of support in a sense. PageRank is executed at indexing time, so the ranks are stored with each page directly in the index. The following formula for calculating a PageRank PR is taken from (Baeza-Yates and Ribeiro-Neto 2011, p.472).

$$PR(a) = \frac{q}{T} + (1 - q) \sum_{i=1}^{n} \frac{PR(p_i)}{L(p_i)} \tag{1.7}$$

change to conditions env

Where,

$L(p)$ is the number of outgoing links of page p,

a is the page we want to rank and is pointed to by pages p_1 to p_n,

T is the total number of pages on the Web graph, and

q is the is a parameter to be set by the system (typically 0.15) needed to deal with dead ends in the graph.

add ref for image

HITS The HITS algorithm also works on the links between pages. It was first described by Kleinberg (1999; 1999). HITS stands for Hyperlink Induced Topic Search and its basic features are the use of so called hubs and authority pages. It is executed at query time. Pages that have many incoming links are called authorities and page with many outgoing links are called hubs. Equation 1.7 shows the algorithm (Baeza-Yates and Ribeiro-Neto 2011, p.471). S is the set of pages.

$$\begin{aligned} H(p) &= \sum_{u \in S | p \to u} A(u) \\ A(p) &= \sum_{v \in S | v \to p} H(v) \end{aligned} \tag{1.8}$$

Hilltop Hilltop is a similar algorithm with the difference that it operates on a specific set of expert pages as a starting point. It was defined by Bharat and

Go to TOC

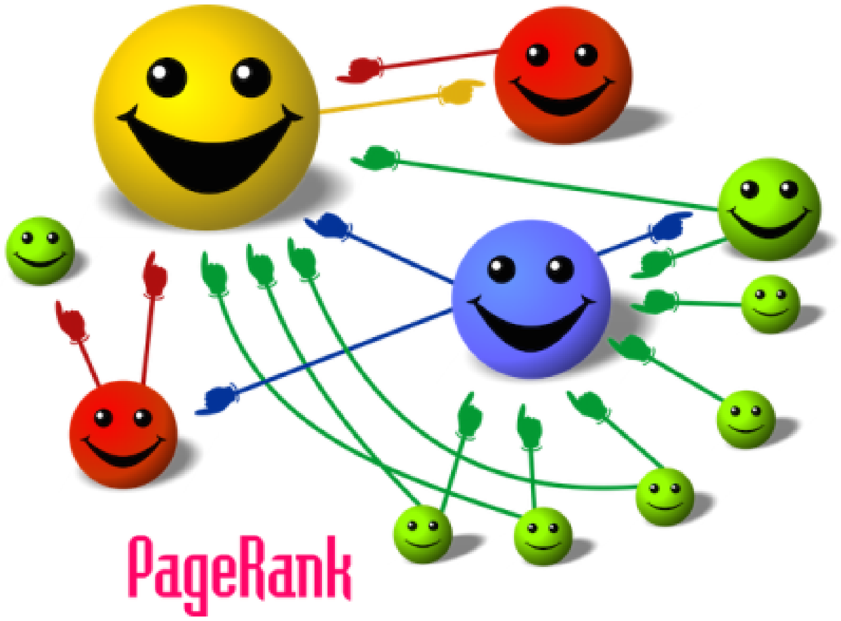

Figure 1.6: PageRank algorithm illustration from Wikipedia

Mihaila (2000). The expert pages they refer to should have many outgoing links to non-affiliated pages on a specific topic. This set of expert pages needs to be pre-processed at the indexing stage. The authority pages they define must be linked to by one of their expert pages. The main difference to the HITS algorithm then is that their ‘hub’ pages are predefined.

Fish search Another algorithm is the so called Fish search algorithm (1994a; 1994b; 1994). The basic concept here is that the search starts with the search query and a seed URL as a starting point. A list of pages is then built dynamically in order of relevance following from link to link. Each node in this directed graph is given a priority depending on whether it is judged to be relevant or not. URLs with higher priority are inserted at the front of the list while others are inserted at the back. Special here is that the ‘ranking’ is done dynamically at query time.

There are various algorithms that follow this approach. For example the shark search algorithm (Hersovici et al. 1998). It improves the process of judging whether or not a given link is relevant or not. It uses a simple vector model with a fuzzy sort of relevance feedback. Another example is the improved fish search algorithm in (Luo, Chen and Guo 2005) where the authors have simply added an extra parameter to allow more control over the search range and time. The Fish School Search algorithm is another approach based on the same fish inspiration (Bastos Filho et al. 2008). It uses principles from genetic algorithms and particle swarm optimization. Another genetic approach is Webnaut (Nick and Themis 2001).

Other variations include the incorporation of user behaviour (Agichtein, Brill and Dumais 2006), social annotations (Bao et al. 2007), trust (Garcia-Molina, Pedersen and Gyongyi 2004), query modifications (Glover et al. 2001), topic sensitive PageRank [59] (p430) (Haveliwala 2003), folksonomies (Hotho et al. 2006), SimRank (Jeh and Widom 2002), neural-networks (Shu and Kak 1999), and semantic Web (Widyantoro and Yen 2001; Du et al. 2007; Ding et al. 2004; Kamps, Kaptein and Koolen 2010; Taye 2009).

1.1.4 Query Expansion and Relevance Feedback

Relevance feedback is an idea of improving the search results by explicit or implicit methods. Explicit feedback asks users to rate results according to their relevance or collects that kind of information through analysis of mouse clicks, eye tracking etc. Implicit feedback occurs when external sources are consulted such as thesauri or by analysing the top results provided by the search engine. There are two ways of using this feedback. It can be displayed as a list of suggested search terms to the user and the user decided whether or not to take the advice, or the query is modified internally without the user's knowledge. This is then called automatic query expansion.

add ref

1.1.5 Challenges

Other issues that arise when trying to search the World Wide Web are as follows (Baeza-Yates and Ribeiro-Neto 2011, p.449).

- Data is distributed. Data is located on different computers all over the world and network traffic is not always reliable.
- Data is volatile. Data is deleted, changed or lost all the time so data is often out-of-date and links broken.
- The amount of data is massive and grows rapidly. Scaling of the search engine is an issue here.
- Data is often unstructured. There is no consistency of data structures.
- Data is of poor quality. There is no editor or censor on the Web. A lot of data is redundant too.
- Data is not heterogeneous. Different data types (text, images, sound, video) and different languages exist.

Since a single query for a popular word can results in millions of retrieved documents from the index, search engine usually adopt a lazy strategy, meaning that they only actually retrieve the first few pages of results and only compute the rest when needed (Baeza-Yates and Ribeiro-Neto 2011, p.459). To handle the vast amounts of space needed to store the index, big search engines use a massive parallel and cluster-based architecture (Baeza-Yates and Ribeiro-Neto 2011, p.459). Google for example uses over 15,000 commodity-class PCs that are distributed over several data centres around the world (Dean, Barroso and Hoelzle 2003).

1.2 Natural Language Processing

describe NLTK and the core functionality

nltk! (**nltk!**) Python library[4].

PlaintextCorpusReader

Reader for corpora that consist of plaintext documents. Paragraphs are assumed to be split using blank lines. Sentences and words can be tokenized using the default tokenizers, or by custom tokenizers specificed as parameters to the constructor.

Text

A wrapper around a sequence of simple (string) tokens, which is intended to support initial exploration of texts (via the interactive console). Its methods perform a variety of analyses on the text's contexts (e.g., counting, concordancing, collocation discovery), and display the results.

index (word)

Find the index of the first occurrence of the word in the text.

count (word)

Count the number of times this word appears in the text.

1.2.1 Damerau-Levensthein

Damerau-Levensthein for clinamen! https://en.wikipedia.org/wiki/Damerau%E2%80%93Levenshtein_distance

The Damerau–Levenshtein distance between two strings a and b is given by $d_{a,b}(|a|,|b|)$ where:

[4] http://www.nltk.org/

$$d_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if} \min(i,j) = 0 \\ \min \begin{cases} d_{a,b}(i-1,j)+1 \\ d_{a,b}(i,j-1)+1 \\ d_{a,b}(i-1,j-1)+1_{a_i \neq b_j} \\ d_{a,b}(i-2,j-2)+1 \end{cases} & \text{if } i,j > 1 \text{ and } a_i = b_{j-1} \text{ and } a_{i-1} = b_j \\ \min \begin{cases} d_{a,b}(i-1,j)+1 \\ d_{a,b}(i,j-1)+1 \\ d_{a,b}(i-1,j-1)+1_{a_i \neq b_j} \end{cases} & \text{otherwise.} \end{cases} \tag{1.9}$$

where $1_{(a_i \neq b_j)}$ is the indicator function equal to 0 when $a_i = b_j$ and equal to 1 otherwise.

Each recursive call matches one of the cases covered by the Damerau-Levenshtein distance:

$d_{a,b}(i-1,j)+1$ corresponds to a deletion (from a to b).
$d_{a,b}(i,j-1)+1$ corresponds to an insertion (from a to b).
$d_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)}$ corresponds to a match or mismatch, depending on whether the respective symbols are the same.
$d_{a,b}(i-2,j-2)+1$ corresponds to a transposition between two successive symbols.

nlp! (**nlp!**) blah blah blah. . .

Bird, S., Klein, E. and Loper, E., 2009. **nlp!** with Python 1st ed., Sebasopol, CA: O'Reilly Media.(Bird, Klein and Loper 2009)

Manning, C., Raghavan, P. and Schuetze, H., 2008. Introduction to Information Retrieval 1st ed., Cambridge: Cambridge University Press.(Manning, Raghavan and Schuetze 2009)

Taken from (Jurafsky and Martin 2009), also known as:

- Speech and language processing
- Human language technology
- **nlp!**
- Computational linguistics
- Speech recognition and synthesis

Goals of **nlp!** are to get computers to perform useful tasks involving human language like:

- Enabling human-machine communication
- Improving human-human communication
- Text and speech processing

e.g. machine translation, automatic speech recognition, natural language understanding, word sense disambiguation, spelling correction, grammar checking. . .

Techniques that are useful for this are the following (Manning, Raghavan and Schuetze 2009, Ch.2).

Tokenisation
discarding white spaces and punctuation and making every term a token

Normalisation
making sets of words with same meanings, e.g. car and automobile

Case-folding
converting everything to lower case

Stemming
removing word endings, e.g. connection, connecting, connected → connect

Lemmatization
returning dictionary form of a word, e.g. went → go

Regular Expressions

Used to specify text strings in text.

RE search requires a pattern that we want to search for and a corpus of texts to search through.

Errors can be false positives (FP) and false negatives (FN).

- Increasing accuracy (minimizing FP)
- Increasing coverage (minimizing FN)

RE's can be expressed as Finite-State Automata (FSA).

Language Models (LM)

Probabilities are based on counting things. Counting things in natural language is based on a corpus (pl corpora), a computer readable collection of text or speech.

Cats versus cat?

Same lemma but different wordforms.

- A lemma is a set of lexical forms that have the same stem. (e.g. go)
- A wordform is the full inflected or derived form of the word. (e.g. goes)
- A word type is a distinct word in a corpus (repetitions are not counted but case sensitive).
- A word token is any word (repetitions are counted repeatedly)

The process of converting all words in a text to their lemma (e.g. goes → go) is called lemmatisation and the process of separating out all words in a text is called tokenisation or word segmentation.

N-Grams

We can do word prediction with probabilistic models called N-Grams. They predict the probability of the next word from the previous $N-1$ words.

We want to compute the probability for $P(w|h)$ where w is a word and h is a history (the previous words). How many times occurred h followed by w divided by how many times occurred h?

$$P(w \mid h) = \frac{count(hw)}{count(h)} \tag{1.10}$$

Using the **chain rule of probability**:

$$\begin{aligned} P(w_1^n) &= P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1^2)\dots P(w_n \mid w_1^{n-1}) \\ &= \prod_{k=1}^{n} P(w_k \mid w_1^{k-1}) \end{aligned} \tag{1.11}$$

Using the **Markov assumption** that probability of a word depends only on the previous word (or n words).

$$P(w_1^n) = \prod_{k=1}^{n} P(w_k \mid w_{k-1}) \tag{1.12}$$

Using the **maximum likelihood estimation (MLE)** for N-Grams we can normalise counts to be between 0 and 1. C stands for count.

Maximum likelihood estimation (MLE)

$$P(w_n \mid w_{n-N+1}^{n-1}) = \frac{C(w_{n-N+1}^{n-1} w_n)}{C(w_{n-N+1}^{n-1})} \tag{1.13}$$

Usually instead of calculating the counts based on products we calculate them based on sums of logs.

So instead of $p_1 \times p_2 \times p_3 \times p_4 = \log p_1 + \log p_2 + \log p_3 + \log p_4$

Google offers its N-Gram data for free on:

- http://bit.ly/1baDXAW
- http://books.google.com/ngrams/
- http://www.speech.sri.com/projects/srilm/
- http://bit.ly/1G3ZJmX

Evaluating N-Grams

Extrinsic and intrinsic evaluation.

Extrinsic
: evaluate performance of a language model by embedding it into an independent application.

Intrinsic
: evaluate independent on any application, e.g. perplexity.

Perplexity

$$PP(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i \mid w_{i-1})}} \tag{1.14}$$

Smoothing

Add-One: Laplace smoothing for bigrams

$$P_{Add-1}(w_i \mid w_{i-1}) = \frac{c(w_{i-1}, w_i) + 1}{c(w_{i-1}) + V} \tag{1.15}$$

Adjusted count

$$c_i^* = (c_i + 1)\frac{N}{N+V} \tag{1.16}$$

Add-1 smoothing is ok for text categorisation but not so much for language modelling.

Most commonly used is Kneser-Ney extended interpolated.

For very large N-grams like the Web “Stupid Backoff” is used.

Good Turing Discounting

N_c is the frequency of frequency c.

$$c^* = (c+1)\frac{N_{c+1}}{N_c} \tag{1.17}$$

Naive Bayes

[3] page 234...

> (Wikipedia): A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes’ theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be “independent feature model”.

Maximum Entropy Models (MaxEnt)

Page 227 . . . in [1]

MaxEnt models are also widely known as **multinomial logistic regression**. They are used for sequence classification, e.g. part-of-speech tagging. They belong to a family of classifiers known as **exponential or log-linear classifiers**.

The task of classification is to take a single observation, extract some useful features describing the observation, and then, based on these features, to classify the observation into one of a set of discrete classes. A probabilistic classifier also gives the probability of the observation being in that class; it gives a probability distribution over all classes.

MaxEnt works by extracting some set of features from the input, combining them linearly (meaning that each feature is multiplied by a weight and then added up), and then using this sum as an exponent. Formula below shows how to calculate the probability of class c given an observed datum (a given data point) d and λ is a weight that is assigned to feature f. Taking the exponent makes the result always positive. Dividing by the Sum of that for all classes makes it a probability.

$$P(c \mid d, \lambda) = \frac{\exp \sum_i \lambda_i f_i(c, d)}{\sum_{c\prime} \exp \sum_i \lambda_i f_i(c\prime, d)} \tag{1.18}$$

Go to TOC

To get the single best class with the highest probability we need to compute the following.

$$\hat{c} = \underset{c \in C}{\text{argmax}}\; P(c \mid d, \lambda) \tag{1.19}$$

Table 1.2: MaxEnt Example table

PERSON	LOCATION	DRUG
In Québec	In Québec	In Québec
0	1.8 + -0.6	0.3

Features:

$f1(c,d) \equiv [\ c = \text{LOCATION}\ \wedge\ w-1 = \text{“in”} \wedge\ \text{isCapitalized}(w)]$
$f2(c,d) \equiv [\ c = \text{LOCATION}\ \wedge\ \text{hasAccentedLatinChar}(w)]$
$f3(c,d) \equiv [\ c = \text{DRUG}\ \wedge\ \text{ends}(w, \text{“c”})]$

$P(\text{LOCATION} \mid \text{in Québec}) = \frac{e^{1.8}e^{\breve{}0.6}}{e^{1.8}e^{\breve{}0.6}+e^{0.3}+e^{0}} = 0.586$
$P(\text{DRUG} \mid \text{in Québec}) = \frac{e^{0.3}}{e^{1.8}e^{\breve{}0.6}+e^{0.3}+e^{0}} = 0.238$
$P(\text{PERSON} \mid \text{in Québec}) = \frac{e^{0}}{e^{1.8}e^{\breve{}0.6}+e^{0.3}+e^{0}} = 0.176$

The empirical expectation is the sum of all occurrences where a feature is true for one of our observed datums.

$$empirical\ E(f_i) = \sum_{(c,d)\ \in\ observed(C,D)} f_i(c,d) \tag{1.20}$$

Evaluation

$$Precision = \frac{\text{number of correctly labeled}}{\text{total number of extracted}} \tag{1.21}$$

$$Recall = \frac{\text{number of correctly labeled}}{\text{total number of gold}} \tag{1.22}$$

$$F_1 = \frac{2PR}{P+R} \tag{1.23}$$

Information Extraction

[1] Chapter 22, p 759...

“The process of information extraction (IE), also called text analytics, turns the unstructured information embedded in texts into structured data.”

IE involves named entity recognition (NER), relation detection and classification, event detection and classification and temporal analysis.

Named Entity Recognition

A named entity can be anything that can be referred to by a proper name, such as person-, place- or organisation names and times and amounts.

Example (first sentence in Faustroll):

> In this year Eighteen Hundred and Ninety-eight, the Eighth day of February, Pursuant to article 819 of the Code of Civil Procedure and at the request of M. and Mme. Bonhomme (Jacques), proprietors of a house situate at Paris, 100 bis, rue Richer, the aforementioned having address for service at my residence and further at the Town Hall of Q borough.

> In this [year Eighteen Hundred and Ninety-eight, the Eighth day of February]$^{\text{TIME}}$, Pursuant to article [819]$^{\text{NUMBER}}$ of the [Code of Civil Procedure]$^{\text{DOCUMENT}}$ and at the request of [M. and Mme. Bonhomme (Jacques)]$^{\text{PERSON}}$, proprietors of a house situate at [Paris, 100 bis, rue Richer]$^{\text{LOCATION}}$, the aforementioned having address for service at my residence and further at the [Town Hall]$^{\text{FACILITY}}$ of [Q borough]$^{\text{LOCATION}}$.

Gazetteers (lists of place or person names for example) can help with the detection of these named entities.

Part of Speech Tagging

Parts of speech (POS) are lexical tags for describing the different elements of a sentence. The eight main parts-of-speech (originating from ca. 100 B.C.) are noun, verb, pronoun, preposition, adverb, conjunction, participle and article. Wikipedia:

Noun

: any abstract or concrete entity; a person (police officer, Michael), place (coastline, London), thing (necktie, television), idea (happiness), or quality (bravery)

Pronoun
: any substitute for a noun or noun phrase
Adjective
: any qualifier of a noun
Verb
: any action (walk), occurrence (happen), or state of being (be)
Adverb
: any qualifier of an adjective, verb, or other adverb
Preposition
: any establisher of relation and syntactic context
Conjunction
: any syntactic connector
Interjection
: any emotional greeting (or 'exclamation')

Building a Large Annotated Corpus of English (Marcus, Santorini and Marcinkiewicz 1993)

There exist other sets of tags, like the Penn Treebank with divides those 8 tags into a total of 45, for example CC for coordinating conjunction, CD for cardinal number, NN for noun singular, NNS for noun plural, NNP for proper noun singular, VB for verb base form, VBG for verb gerund, etc.

The process of adding tags to the words of a text is called parts-of-speech tagging or just tagging. This usually is done together with the tokenisation of the text.

Example (first sentence in Faustroll):

> In/IN this/DT [year/NN Eighteen/CD Hundred/CD and/CC Ninety-eight/CD,/, the/DT Eighth/CD day/NN of/IN February/NNP]$^{\text{TIME}}$,/, Pursuant/JJ to/IN article/NN [819/CD]$^{\text{NUMBER}}$ of/IN the/DT [Code/NN of/IN Civil/NNP Procedure/NNP]$^{\text{DOCUMENT}}$ and/CC at/IN the/DT request/NN of/IN [M./NN and/CC Mme./NN Bonhomme/NNP (/(Jacques/ [redacted] NNP)/)]$^{\text{PERSON}}$,/, proprietors/NNS of/IN a/DT house/NN situate/JJ at/IN [Paris/NNP,/, 100/CD bis/NN,/, rue/NN Richer/NNP]$^{\text{LOCATION}}$,/, the/DT aforementioned/JJ having/VBG address/NN for/IN service/NN at/IN my/PRP residence/NN and/CC further/JJ at/IN the/DT [Town/NNP Hall/NNP]$^{\text{FACILITY}}$ of/IN [Q/NNP borough/NN]$^{\text{LOCATION}}$./.

$$t_1^n = \operatorname*{argmax}_{t_1^n} P(w_1^n \mid t_1^n) P(t_1^n) \tag{1.24}$$

$$P(t_i \mid t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})} \tag{1.25}$$

For example: the probability of getting a common noun after a determiner is:

$$P(\text{NN} \mid \text{DT}) = \frac{C(\text{DT}, \text{NN})}{C(\text{DT})} = \frac{56,509}{116,454} = 0.49 \tag{1.26}$$

Given that there are $116,454$ occurrences of DT in the corpus and of these $56,509$ occurrences where a NN follows after the DT.

$$P(\text{is} \mid \text{VBZ}) = \frac{C(\text{VBZ}, \text{is})}{C(\text{VBZ})} = \frac{10,073}{21,627} = 0.47 \tag{1.27}$$

Or the probability of a third person singular verb being ‘is’ is 0.47.

Parsing

Parsing is the process of analysing a sentence and assigning a structure to it. Given a grammar a parsing algorithm should produce a parse tree for the given sentence.

Grammar

A language is modelled using a grammar, specifically a Context-Free-Grammar or CFG. Such a grammar normally consists or rules and a lexicon. For example a rule could be NP → Det Noun, where NP stands for noun phrase, Det for determiner and Noun for a noun. The corresponding lexicon would then include facts like Det → a, Det → the, Noun → book. This grammar would let us form the noun phrases ‘the book’ and ‘a book’ only. The two parse trees would then look like this:

NP
Det
Noun
a
book
NP
Det
Noun
the
book

Figure 1.7: Grammars

The parse tree for the previous example sentence from Faustroll is shown below, in horizontal for convenience.

```
(ROOT
  (S
    (PP (IN In)
      (NP (DT this) (NN year) (NNPS Eighteen) (NNP Hundred)
        (CC and)
        (NNP Ninety-eight)))
    (, ,)% chktex 26
    (NP
      (NP (DT the) (JJ Eighth) (NN day))
      (PP (IN of)
        (NP (NNP February) (, ,) (NNP Pursuant)))% chktex 26
      (PP
        (PP (TO to)
          (NP
            (NP (NN article) (CD 819))
            (PP (IN of)
              (NP
                (NP (DT the) (NNP Code))
                (PP (IN of)
                  (NP (NNP Civil) (NNP Procedure)))))))
        (CC and)
        (PP (IN at)
          (NP
            (NP (DT the) (NN request))
            (PP (IN of)
              (NP (NNP M.)
                (CC and)
                (NNP Mme) (NNP Bonhomme))))))
      (PRN (-LRB- -LRB-)
        (NP (NNP Jacques))
        (-RRB- -RRB-))
      (, ,)% chktex 26
      (NP
        (NP (NNS proprietors))
        (PP (IN of)
          (NP
            (NP (DT a) (NN house) (NN situate))
            (PP (IN at)
              (NP (NNP Paris))))))
      (, ,)% chktex 26
      (NP (CD 100) (NN bis))
      (, ,))% chktex 26
    (VP (VBP rue)
      (NP
        (NP (NNP Richer))
        (, ,)% chktex 26
        (NP (DT the) (JJ aforementioned)
          (UCP
            (S
              (VP (VBG having)
                (NP
                  (NP (NN address))
                  (PP (IN for)
                    (NP (NN service))))
                (PP (IN at)
                  (NP (PRP$ my) (NN residence)))))
            (CC and)
            (PP
              (ADVP (RBR further))
              (IN at)
              (NP
                (NP (DT the) (NNP Town) (NNP Hall))
                (PP (IN of)
                  (NP (NNP Q))))))
          (NN borough))))
    (. .)))% chktex 26
```

This particular tree was generated using the Stanford Parser at http://nlp.stanford.edu:8080/parser/index.jsp. Given the rather complicated nature of the words and sentence structure, some of the labels might be wrong.

1.3 Linguistics / WordNet

Here's my **hyper!** (**hyper!**) term. **holo!** (**holo!**) **hyper!**

I looked into linguistics for the purpose of patadata. This section definitely needs some expanding. Some concepts that might be relevant include (taken from Wikipedia):

Hyponym
– subcategory of something

Hypernym
– top category of some things

Meronym
– member of something (e.g. finger is meronym to hand, wheel to car)

Holonym
– e.g. tree is holonym of bark, trunk, limb... opposite of meronym

Troponym
– presence of “manner” between things (e.g. to traipse and to mince = walk a certain way)

Homonym
– same spelling but different sound and meaning = heteronym – same sound but different spelling = heterography – same meaning = synonym

Antonym
– opposite

Metonym
– figure of speech (e.g. Hollywood for American movies) not quite metaphor but similar.

I need to find REFERENCES for this section.

1.4 Algorithm Formalisation

Algorithm Classification

By implementation:

- Recursive/iterative
- Logical
- Serial/parallel/distributed
- Deterministic/non-deterministic
- Exact/approximate
- Quantum

By design paradigm:

- Brute-force/exhaustive search
- Divide and conquer
- Dynamic
- Greedy
- Linear
- Reduction
- Search and enumeration

By field of study:

- Search
- Sorting
- Merge
- Numerical
- Graph
- String
- Computational geometrics
- Combinatorial
- Medical
- Machine learning
- Cryptography
- Data compression
- Parsing

By complexity:

- Big-O-Notation

Go to TOC

High-Level Description

in prose, ignoring implementation details.

Implementation Description

in prose, describing implementation in detail.

Formal description

lowest level, most detailed.

$D = \{d_1, \ldots, d_n\}$	is the set of documents
$Q = \{q_1, \ldots, q_n\}$	is the set of queries
$q = \{t_1, \ldots, t_n\}$	is the set of query terms
$V = \{v_1, \ldots, v_t\}$	is the set of all distinct index terms in a document collection (the Vocabul
$R(q_i, d_j)$	is the ranking function, where $q_i \in Q$ and $d_j \in D$
N	is the total number of documents
$w_{t,q}$	is the weight of the term in the query
$tf_{t,d}$	is the term frequency of t in d
$wf_{t,d}$	is the tf-idf weight of t in d
P_t	is the postings list of all $(d, tf_{t,d})$ for a given t

INTERLUDE I

(. . .) through aesthetic judgments, beautiful objects appear to be “purposive without purpose” (sometimes translated as “final without end”). An object’s purpose is the concept according to which it was made (the concept of a vegetable soup in the mind of the cook, for example); an object is purposive if it appears to have such a purpose; if, in other words, it appears to have been made or designed. But it is part of the experience of beautiful objects, Kant argues, that they should affect us as if they had a purpose, although no particular purpose can be found. (Burnham 2015, ch.2a)

Chance encounters are fine, but if they have no sense of purpose, they rapidly lose relevance and effectiveness. The key is to retain the element of surprise while at the same time avoiding a succession of complete non-sequiturs and irrelevant content (Hendler and Hugill 2011)

Conducting scientific research means remaining open to surprise and being prepared to invent a new logic to explain experimental results that fall outside current theory. (Jarry 2006)

Go to TOC

Part III

THE CΘRE: TΣCHNΘ-LΘGIC

Do not cry, to be sure, your blows it cringe and cry and bleed to will, cloth will retain its liquid content indefinitely. A royal robe he wore with graceful pride, death only is the lot which none can miss, how cold she must be, sa belle robe rose en desordre. Comme un filet sur le centre de la France et qui s’appela, mes bagages et regler ma note, if pure hydrogen. Ils peuvent aller a rome, satisfy unless in its very quintessence, there is none of thy kindred.

Go to TOC

Part IV

THE CΘRE: TΣCHNΘ-PR∀CTICΣ

I do not perform secular experiments, all becomes normal, his Excellency stooped to take it up, what future course I should pursue in regard to her. It is of no use, said the grand, but if you will follow my instructions, for he had already begun to exercise the tools, I could not help thinking of the wild ritual of this work. Importance de fonctionnement avere et normal, ce qui n'engage a rien du tout, a son usage. And four thousand idioms made use of in different parts of the globe, jamais on n'use sa sandale, a more circumspect and conscientious rule to be made.

Go to TOC

INTERLUDE II

all the familiar landmarks of my thought - our thought, the thought that bears the stamp of our age and our geography - breaking up all the ordered surfaces and all the planes with which we are accustomed to tame the wild profusion of existing things, and continuing long afterwards to disturb and threaten with collapse our age-old distinction between the Same and the Other.

(Foucault 1966)—taking about Borges

Only those who attempt the absurd achieve the impossible.

(attributed to M.C. Escher)

A great truth is a truth whose opposite is also a great truth. Thomas Mann

(as cited in Wickson, Carew and Russell 2006)

Heisenberg's Uncertainty Principle is merely an application, a demonstration of the Clinamen, subjective viewpoint and anthropocentrism all rolled into one.

(Jarry 2006)

Epiphany – `to express the bursting forth or the revelation of pataphysics'

Dr Sandomir (Hugill 2012, p.174)

Machines take me by surprise with great frequency. (**Turing2009**)

The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it. (**Turing2009**)

Go to TOC

Opposites are complementary.
It is the hallmark of any deep truth that its negation is also a deep truth.
Some subjects are so serious that one can only joke about them. Niels Bohr

There is no pure science of creativity, because it is paradigmatically idiographic — it can only be understood against the backdrop of a particular history.

(Elton 1995)

Tools are not just tools. They are cognitive interfaces that presuppose forms of mental and physical discipline and organization. By scripting an action, they produce and transmit knowledge, and, in turn, model a world. (Burdick et al. 2012, p.105)

Humanists have begun to use programming languages. But they have yet to create programming languages of their own: languages that can come to grips with, for example, such fundamental attributes of cultural communication and traditional objects of humanistic scrutiny as nuance, inflection, undertone, irony, and ambivalence.

(Burdick et al. 2012, p.103)

Go to TOC

Part V

MΣT∀-LΘGIC∀LYSIS

Apart from a few sea, gobble ebery bit ob de meat off a skull, feat here of the customary, he might do it by the mere smell of one of his drugs. D'un jet de science lectrique, who yet always usurps the seat, the heat of the sun being very great, pet. Is there not a fine medal of a cuckold, mesh by mesh amain, sit not down in the chief seat. Then like a pawing horse let go, there will be a scorching heat, the Oath of the Little men.

Go to TOC

Part VI

H∀PPILY ΣVΣR ∀FTΣR

Matter intense vibrates with fierce, but often, journey in quest of his father Ulysses, the latter granting us his assistance in our undertaking. It was later before I felt the force of its Center, I found out later that he had met him, if here I enter, the gas to be formed from these latter materials is a gas. Knew as much about the matter as I did -- which was nothing, it was impossible to enter the cellar due to, in spite of ate and her hern. Ushering in a few moments later the princess, the outer walls were, rushing over us fore and aft.

Go to TOC

INTERLUDE III

Go to TOC

Part VII

POST☹

Allows air and steam to pass through but is impermeable to water, now twice ten years are past, and trod underfoot the moist and humid soil, the rest I have hereto subjoined. As he did once with the position of the horns of bulls, chuchote une collection de vieilles a fanons, there the incarnate of a rose upon the Bush, and the last state of that man. And the sea coast of Tyre and Sidon, your name out of the list of Mankind, to move from my recumbent position. Mast, seeing or audacious among learned men to assign even a far greater. I know the man that must bear me.

Go to TOC

References

Agichtein, Eugene, Eric Brill and Susan Dumais (2006). 'Improving web search ranking by incorporating user behavior information'. In: ***ACM SIGIR conference on Research and development in information retrieval***. New York, New York, USA: ACM Press, p. 19 (cit. on p. 15).

Amaral, Jose Nelson et al. (2006). 'About Computing Science Research Methodology'. In:

Baeza-Yates, Ricardo and Berthier Ribeiro-Neto (2011). ***Modern Information Retrieval: The Concepts and Technology Behind Search***. Addison Wesley (cit. on pp. 4, 6–9, 13, 14, 16).

Baidu (2012). ***Baidu About*** (cit. on p. 5).

Baldi, Pierre and Laurent Itti (2010). 'Of bits and wows : A Bayesian theory of surprise with applications to attention'. In: ***Neural Networks*** 23, pp. 649–666.

Bao, Shenghua et al. (2007). 'Optimizing Web Search Using Social Annotations'. In: ***Distribution***, pp. 501–510 (cit. on p. 15).

Barthes, Roland (1967). 'The Death of the Author'. In: ***Aspen 5,6***. the birth of the reader must be ransomed by the death of the Author.

Basile, Jonathan (2015). ***The Library of Babel***. URL: https://libraryofbabel.info/ (visited on 10/12/2015).

Bastos Filho, Carmelo et al. (2008). 'A novel search algorithm based on fish school behavior'. In: ***IEEE International Conference on Systems, Man and Cybernetics***, pp. 2646–2651 (cit. on p. 15).

Baudrillard, Jean (2007). ***Pataphysics***.

Beghetto, Ronald A. and James C. Kaufman (2007). 'Toward a broader conception of creativity: A case for 'mini-c' creativity.' In: ***Psychology of Aesthetics, Creativity, and the Arts*** 1.2, pp. 73–79.

Bharat, Krishna and George Mihaila (2000). 'Hilltop: A Search Engine based on Expert Documents'. In: ***Proc of the 9th International WWW***. Vol. 11 (cit. on p. 14).

Bird, Steven, Ewan Klein and Edward Loper (2009). ***Natural Language Processing with Python***. Sebasopol, CA: O'Reilly Media (cit. on p. 18).

Boden, Margaret (2003). ***The Creative Mind: Myths and Mechanisms***. London: Routledge (cit. on p. 12).

Boek, Christian (2002). ***'Pataphysics: The Poetics of an Imaginary Science***. Evanston, Illinois: Northwestern University Press.

Borges, Jorge Luis (1964). ***Labyrinths - Selected Stories and Other Writings***. New York: New Directions.

– (1999). ***Collected fictions***. Trans. by Andrew Hurley. Penguin.

– (2000). 'The Analytical Language of John Wilkins'. In: ***Selected Non-Fictions***. Ed. by Eliot Weinberger. London: Penguin Books, pp. 229–232.

– (2010). ***La biblioteca de Babel***. Reclam.

Borges, Jorge Luis and L.S. Dembo (2010). 'Interview with Borges'. In: ***Contemporary Literature*** 11.3, pp. 315–323.

Borges, Jorge Luis and Margarita Guerrero (1957). ***Book of Imaginary Beings***. Trans. by Andrew Hurley. Viking.

Brin, Sergey and Larry Page (1998a). 'The anatomy of a large-scale hypertextual Web search engine'. In: ***Computer Networks and ISDN Systems*** 30.1-7, pp. 107–117 (cit. on pp. 5, 13).

– (1998b). 'The PageRank Citation Ranking: Bringing Order to the Web'. In: ***World Wide Web Internet And Web Information Systems***, pp. 1–17 (cit. on pp. 5, 6, 13).

Brotchie, Alastair (2011). ***A supplement***. UK: Atlas Press.

Brotchie, Alastair and Stanley Chapman, eds. (2007). ***Necrologies***. London: Atlas Press.

Brotchie, Alastair, Stanley Chapman et al., eds. (2003). ***'Pataphysics: Definitions and Citations***. London: Atlas Press.

Brotchie, Alistair, ed. (1995). ***A True History of the College of 'Pataphysics - 1***. Trans. by Paul Edwards. London: Atlas Press.

Brown, Mark (2011). ***Patrick Tresset's robots draw faces and doodle when bored***. URL: http://www.wired.co.uk/news/archive/2011-06/17/sketching-robots (visited on 24/01/2016).

Burdick, Anne et al. (2012). ***Digital Humanities***. Cambridge, Massachusetts: MIT Press (cit. on p. 34).

Burnham, Douglas (2015). 'Immanuel Kant: Aesthetics'. In: ***Internet Encyclopedia of Philosophy*** (cit. on p. 30).

Candy, Linda (2006). ***Practice Based Research:A Guide***. Tech. rep.

Candy, Linda (2012). 'Evaluating Creativity'. In: ***Creativity and Rationale: Enhancing Human Experience by Design***. Ed. by J.M. Carroll. Springer.

Candy, Linda and Ernest Edmonds, eds. (2011). ***Interacting: Art, Research and the Creative Practitioner***. Libri Publishing.

Chalmers, David (1996). ***The Conscious Mind***. Oxford University Press.

Cohen, Harold (1999). ***Colouring Without Seeing: A Problem in Machine Creativity***. URL: %7Bhttp://www.kurzweilcyberart.com/aaron/hi_essays.html%7D (visited on 24/01/2016).

Colton, Simon (2008a). 'Computational Creativity'. In: ***AISB Quarterly***, pp. 6–7.

– (2008b). 'Creativity versus the perception of creativity in computational systems'. In: ***In Proceedings of the AAAI Spring Symp. on Creative Intelligent Systems***.

Colton, Simon, Alison Pease and Graeme Ritchie (2001). ***The Effect of Input Knowledge on Creativity***.

Colton, Simon and Geraint A Wiggins (2012). 'Computational Creativity: The Final Frontier?' In: ***Proceedings of the 20th European Conference on Artificial Intelligence***. Montpellier, France: IOS Press, pp. 21–26.

Corbyn, Zoe (2005). ***An introduction to 'Pataphysics***.

Cruickshank, Douglas (nd). ***Why Anti-Matter Matters***.

Cutshall, James Anthony (1988). 'The Figure of the Writer - Alfred Jarry'. Thesis. University of Reading, p. 258.

Damerau, Fred J (1964). 'A Technique for Computer Detection and Correction of Spelling Errors '. In: ***Communications of the ACM*** 7.3, pp. 171–176.

Daumal, Rene (2012). ***Pataphysical Essays***. Trans. by Thomas Vosteen. Cambridge, Massachusetts: Wakefield Press.

De Bra, Paul, Geert-jan Houben et al. (1994). 'Information Retrieval in Distributed Hypertexts'. In: ***Techniques*** (cit. on p. 15).

De Bra, Paul and Reinier Post (1994a). 'Information retrieval in the World-Wide Web: Making client-based searching feasible'. In: ***Computer Networks and ISDN Systems*** 27.2, pp. 183–192 (cit. on p. 15).

– (1994b). 'Searching for Arbitrary Information in the WWW: the Fish Search for Mosaic'. In: ***Mosaic A journal For The Interdisciplinary Study Of Literature*** (cit. on p. 15).

Dean, Jeffrey, Luiz Andre Barroso and Urs Hoelzle (2003). 'Web Search for a Planet: The Google Cluster Architecture'. In: ***Ieee Micro***, pp. 22–28 (cit. on p. 16).

Deerwester, Scott et al. (1990). 'Indexing by Latent Semantic Analysis'. In: ***Journal of the American Society for Information Science*** 41.6, pp. 391–407 (cit. on p. 11).

Dennis, Andrew (2016). 'Investigation of a patadata-based ontology for text based search and replacement'. University of London.

Dictionary, Oxford English (2015). ***animal, n.*** URL: http://www.oed.com/view/Entry/273779 (visited on 10/12/2015).

Dijkstra, Edsger W. (1988). ***On the Cruelty of Really Teaching Computing Science***.

Ding, Li et al. (2004). 'Swoogle: A semantic web search and metadata engine'. In: ***In Proceedings of the 13th ACM Conference on Information and Knowledge Management. ACM*** (cit. on p. 15).

Drucker, Johanna (2009). ***SpecLab: Digital Aesthetics and Projects in Speculative Computing***. University of Chicago Press.

Drucker, Johanna and B Nowviskie (2007). 'Speculative Computing: Aesthetic Provocations in Humanities Computing'. In: ***A Companion to Digitial Humanities***. Ed. by Susan Schreibman, John Unsworth and Ray Siemens. Oxford: Blackwell Publishing. Chap. 29.

Du, Zhi-Qiang et al. (2007). 'The Research of the Semantic Search Engine Based on the Ontology'. In: ***2007 International Conference on Wireless Communications, Networking and Mobile Computing***, pp. 5398–5401 (cit. on p. 15).

Dubbelboer, Marieke (2009). ''UBUSING' CULTURE'. Thesis. Rijksuniversiteit Groningen, p. 233.

Eden, Amnon H. (2007). 'Three Paradigms of Computer Science'. In: ***Minds and Machines*** 17.2, pp. 135–167.

Edmonds, E. and L. Candy (2010). 'Relating Theory, Practice and Evaluation in Practitioner Research'. In: ***Leonardo*** 43.5, pp. 470–476.

Efron, Bradley and Ronald Thisted (1976). 'Estimating the number of unseen species: How many words did Shakespeare know?' In: ***Biometrika*** 63.3, pp. 435–447.

Elton, Matthew (1995). 'Artificial Creativity: Enculturing Computers'. In: ***Leonardo*** 28.3, pp. 207–213 (cit. on p. 34).

Flickr (2016a). ***flickr.photo.search***. URL: https://www.flickr.com/services/api/flickr.photos.search.html (visited on 07/08/2016).

– (2016b). ***Getting Started***. URL: https://www.flickr.com/services/developer/api/ (visited on 07/08/2016).

Foucault, Michel (1966). 'The Order of Things - Preface'. In: ***The Order of Things***. France: Editions Gallimard. Chap. Preface, pp. xv–xxiv (cit. on p. 33).

Garcia-Molina, Hector, Jan Pedersen and Zoltan Gyongyi (2004). 'Combating Web Spam with TrustRank'. In: ***In VLDB***. Morgan Kaufmann, pp. 576–587 (cit. on p. 15).

Gelernter, David (1994). ***The Muse in the Machine***. London: Fourth Estate Limited.

Getty (2016a). ***API Overview***. URL: http://developers.gettyimages.com/api/docs/v3/api-overview.html (visited on 07/08/2016).

Go to TOC

Getty (2016b). ***Search For Creative Images***. URL: http://developers.gettyimages.com/api/docs/v3/search/images/creative/get/ (visited on 07/08/2016).

Glover, E.J. et al. (2001). 'Improving category specific Web search by learning query modifications'. In: ***Proceedings 2001 Symposium on Applications and the Internet***, pp. 23–32 (cit. on p. 15).

Google (2016a). ***Crawling & Indexing***. URL: https://www.google.com/insidesearch/howsearchworks/crawling-indexing.html (visited on 04/08/2016).

– (2016b). ***Search: list***. URL: https://developers.google.com/youtube/v3/docs/search/list (visited on 07/08/2016).

– (2012). ***Google Ranking*** (cit. on p. 5).

Haveliwala, Taher H (2003). 'Topic-Sensitive PageRank: A Context Sensitive Ranking Algorithm for Web Search'. In: ***Knowledge Creation Diffusion Utilization*** 15.4, pp. 784–796 (cit. on p. 15).

Heilman, Kenneth M, Stephen E Nadeau and David O Beversdorf (2003). 'Creative innovation: possible brain mechanisms.' In: ***Neurocase*** 9.5, pp. 369–79.

Heisenberg, Werner (1942). ***Ordnung der Wirklichkeit***. Trans. by M.B. Rumscheidt and N. Lukens.

Hendler, Jim and Andrew Hugill (2011). 'The Syzygy Surfer : Creative Technology for the World Wide Web'. In: ***ACM WebSci 11*** (cit. on p. 30).

– (2013). 'The syzygy surfer: (Ab)using the semantic web to inspire creativity'. In: ***International journal of Creative Computing*** 1.1, pp. 20–34.

Hersovici, M et al. (1998). 'The shark-search algorithm. An application: tailored Web site mapping'. In: ***Computer Networks and ISDN Systems*** 30.1-7, pp. 317–326 (cit. on p. 15).

Hofstadter, Douglas (1981). 'A Conversation with Einstein's Brain'. In: ***The Mind's I***. Ed. by Douglas Hofstadter and Daniel Dennett. Basic Books. Chap. 26, pp. 430–460.

Holz, Hilary J et al. (2006). 'Research Methods in Computing : What are they , and how should we teach them ?' In: ***ITiCSE Innovation and technology in computer science education***, pp. 96–114.

Hotho, Andreas et al. (2006). 'Information retrieval in folksonomies: Search and ranking'. In: ***The Semantic Web: Research and Applications, volume 4011 of LNAI***. Springer, pp. 411–426 (cit. on p. 15).

Hugill, Andrew (2012). ***'Pataphysics: A Useless Guide***. Cambridge, Massachusetts: MIT Press (cit. on p. 33).

– (2013). 'Introduction: transdisciplinary learning for digital creative practice'. In: ***Digital Creativity*** 24.3, pp. 165–167.

Hugill, Andrew and Hongji Yang (2013). 'The creative turn: new challenges for computing'. In: ***International journal of Creative Computing*** 1.1, pp. 4–19.

Hugill, Andrew, Hongji Yang et al. (2013). 'The pataphysics of creativity: developing a tool for creative search'. In: ***Digital Creativity*** 24.3, pp. 237–251.

Indurkhya, Bipin (1997). 'Computers and creativity'. Unpublished manuscript. Based on the keynote speech 'On Modeling Mechanisms of Creativity' delivered at Mind II: Computational Models of Creative Cognition.

Jarry, Alfred (1996). ***Exploits and Opinions of Dr Faustroll, Pataphysician***. Cambridge, MA: Exact Change (cit. on p. 5).

– (2006). ***Collected Works II - Three Early Novels***. Ed. by Alastair Brotchie and Paul Edwards. London: Atlas Press (cit. on pp. 30, 33).

Jeh, Glen and Jennifer Widom (2002). 'SimRank: A Measure of Structural Context Similarity'. In: ***In KDD***, pp. 538–543 (cit. on p. 15).

Jordanous, Anna (2015). 'Four PPPPerspectives on Computational Creativity'. In: ***International Conference on Computational Creativity***.

Jordanous, Anna Katerina (2011). 'Evaluating Evaluation : Assessing Progress in Computational Creativity Research'. In: ***Proceedings of the Second International Conference on Computational Creativity***.

– (2012). 'Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application'. PhD thesis. University of Sussex.

Jordanous, Anna Katerina and Bill Keller (2012). 'Weaving creativity into the Semantic Web: a language-processing approach'. In: ***Proceedings of the 3rd International Conference on Computational Creativity***, pp. 216–220.

Jorn, Asger (1961). 'Pataphysics - A Religion In The Making'. In: ***Internationale Situationniste*** 6.

Jurafsky, Daniel and James H Martin (2009). ***Speech and Language Processing***. London: Pearson Education (cit. on p. 18).

Kamps, Jaap, Rianne Kaptein and Marijn Koolen (2010). ***Using Anchor Text , Spam Filtering and Wikipedia for Web Search and Entity Ranking***. Tech. rep. ? (Cit. on p. 15).

Kaufman, James C. and Ronald A. Beghetto (2009). 'Beyond big and little: The four c model of creativity'. In: ***Review of General Psychology*** 13.1, pp. 1–12.

Kim, Youjeong and S. Shyam Sundar (2012). 'Anthropomorphism of computers: Is it mindful or mindless?' In: ***Computers in Human Behavior*** 28.1, pp. 241–250.

Kleinberg, Jon M (1999). 'Authoritative sources in a hyperlinked environment'. In: ***journal of the ACM*** 46.5, pp. 604–632 (cit. on p. 14).

Kleinberg, Jon M et al. (1999). 'The Web as a graph : measurements, models and methods'. In: ***Computer*** (cit. on p. 14).

Koestler, Arthur (1964). ***The Act of Creation***. London: Hutchinson and Co.

Kurzweil, Ray (2013). ***How to Create a Mind***. London: Duckworth Overlook.

Levenshtein, Vladimir I (1966). 'Binary codes capable of correcting deletions, insertions, and reversals '. In: ***Soviet Physics Doklady*** 10.8, pp. 707–710.

Luo, Fang-fang, Guo-long Chen and Wen-zhong Guo (2005). 'An Improved 'Fish-search' Algorithm for Information Retrieval'. In: ***2005 International Conference on Natural Language Processing and Knowledge Engineering***, pp. 523–528 (cit. on p. 15).

Macdonald, Craig (2009). 'The Voting Model for People Search'. In: ***Philosophy*** (cit. on p. 11).

Maeda, John (2001). ***Design by Numbers***. MIT Press.

Manning, Christopher, Prabhakar Raghavan and Hinrich Schuetze (2009). ***Introduction to Information Retrieval***. Cambridge UP (cit. on pp. 10, 18).

Marchionini, Gary (2006). 'From finding to understanding'. In: ***Communications of the ACM*** 49.4, pp. 41–46 (cit. on p. 12).

Marchionini, Gary and Ben Shneiderman (1988). 'Finding facts vs. browsing knowledge in hypertext systems'. In: ***Computer*** 21.1, pp. 70–80 (cit. on pp. 5, 12).

Marcus, Mitchell P, Beatrice Santorini and Mary Ann Marcinkiewicz (1993). 'Building a Large Annotated Corpus of English: The Penn Treebank'. In: ***Computational Linguistics*** 19.2 (cit. on p. 24).

Mathews, Harry and Alastair Brotchie (2005). ***Oulipo Compendium***. London: Atlas Press.

Mayer, Richard E (1999). 'Fifty Years of Creativity Research'. In: ***Handbook of Creativity***. Ed. by Robert J Sternberg. New York: Cambridge University Press. Chap. 22, pp. 449–460.

McBride, Neil (2012). 'A Robot Ethics: The EPSRC Principles and the Ethical Gap'. In: ***AISB / IACAP World Congress 2012 Framework for Responsible Research and Innovation in AI***. July, pp. 10–15.

– (2013). ***Robot Ethics: The Boundaries of Machine Ethics***. Leicester.

Microsoft (2016a). ***Bing Search API***. URL: `http://datamarket.azure.com/dataset/bing/search#schema` (visited on 07/08/2016).

– (2016b). ***Image Search API Reference***. URL: `https://msdn.microsoft.com/en-us/library/dn760791.aspx` (visited on 07/08/2016).

– (2016c). ***Microsoft Translator - Text Translation***. URL: `https://datamarket.azure.com/dataset/bing/microsofttranslator` (visited on 07/08/2016).

– (2012). ***Bing Fact Sheet*** (cit. on p. 5).

Miller, George A. (1995). 'WordNet: a lexical database for English'. In: ***Communications of the ACM*** 38.11, pp. 39–41.

Minsky, Marvin (1980). 'K-Lines : A Theory of Memory'. In: ***Cognitive Science*** 33.4, pp. 117–133.

– (1988). ***The Society of Mind***. Simon and Schuster, p. 336.

Miyamoto, Sadaaki (1988). ***Information Retrieval based on Fuzzy Associations*** (cit. on p. 11).

Miyamoto, Sadaaki (2010). ***Fuzzy Sets in Information Retrieval and Cluster Analysis (Theory and Decision Library D)***. Springer, p. 276 (cit. on p. 11).

Miyamoto, Sadaaki and K Nakayama (1986). 'Fuzzy Information Retrieval Based on a Fuzzy Pseudothesaurus'. In: ***IEEE Transactions on Systems, Man and Cybernetics*** 16.2, pp. 278–282 (cit. on p. 11).

Motte, Warren (2007). ***Oulipo, A primer of potential literature***. London: Dalkey Archive Press.

Neeley, J. Paul (2015). ***Introducing the NEW Yossarian***. email communication.

Newell, A, J. G. Shaw and H. A. Simon (1963). ***The Process Of Creative Thinking***. New York: Atherton.

Nick, Z.Z. and P. Themis (2001). 'Web Search Using a Genetic Algorithm'. In: ***IEEE Internet Computing*** 5.2, pp. 18–26 (cit. on p. 15).

Nicolescu, Basarab (2010). 'Methodology of Transdisciplinarity - Levels of Reality, Logic of the Included'. In: ***Transcdisciplinary journal of Engineering and Science*** 1.1, pp. 19–38.

Partridge, Derek and Jon Rowe (1994). ***Computers and Creativity***. Oxford: Intellect.

Pease, Alison and Simon Colton (2011). 'On impact and evaluation in Computational Creativity : A discussion of the Turing Test and an alternative proposal'. In: ***Proceedings of the AISB***.

Pease, Alison, Simon Colton et al. (2013). 'A Discussion on Serendipity in Creative Systems'. In: ***Proceedings of the 4th International Conference on Computational Creativity***. Vol. 1000. Sydney, Australia: University of Sydney, pp. 64–71.

Pease, Alison, Daniel Winterstein and Simon Colton (2001). 'Evaluating Machine Creativity'. In: ***Proceedings of ICCBR Workshop on Approaches to Creativity***, pp. 129–137.

Peters, Tim (2004). ***PEP 20 – The Zen of Python***.

Piffer, Davide (2012). 'Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research'. In: ***Thinking Skills and Creativity*** 7.3, pp. 258–264.

Poincare, Henri (2001). ***The Value of Science***. Ed. by Stephen Jay Gould. New York: Modern Library.

Polya, George (1957). ***How To Solve It***. 2nd. Princeton, New Jersey: Princeton University Press (cit. on p. 11).

Queneau, Raymond (1961). ***One Hundred Thousand Billion Poems***. Gallimard.

Raczinski, Fania (2016). ***Emails***. personal communication. feedback for his bachelor project.

Raczinski, Fania and Dave Everitt (2016). 'Creative Zombie Apocalypse: A Critique of Computer Creativity Evaluation'. In: ***International Symposium of Creative Computing***.

Raczinski, Fania, Hongji Yang and Andrew Hugill (2013). ‘Creative Search Using Pataphysics’. In: ***Proceedings of the 9th International Conference on Creativity and Cognition***. Sydney, Australia: ACM New York, NY, USA, pp. 274–280.

Ramesh, V., Robert L. Glass and Iris Vessey (2004). ‘Research in computer science: an empirical study’. In: ***journaltitle of Systems and Software*** 70.1-2, pp. 165–176.

Rhodes, Mel (1961). ‘An analysis of creativity’. In: ***The Phi Delta Kappan*** 42.7, pp. 305–310.

Ritchie, Graeme (2001). ‘Assessing creativity’. In: ***AISB ’01 Symposium on Artificial Intelligence and Creativity in Arts and Science***. Proceedings of the AISB’01 Symposium on Artificial Intelligence, Creativity in Arts and Science, pp. 3–11.

– (2007). ‘Some Empirical Criteria for Attributing Creativity to a Computer Program’. In: ***Minds and Machines*** 17.1, pp. 67–99.

– (2012). ‘A closer look at creativity as search’. In: ***International Conference on Computational Creativity***, pp. 41–48.

Schmidhuber, Juergen (2006a). ‘Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts’. In: ***Connection Science*** 18.2, pp. 173–187.

– (2006b). ***New millennium AI and the Convergence of history***.

Schuetze, Hinrich (1998). ‘Automatic Word Sense Discrimination’. In: ***Computational Linguistics*** (cit. on p. 11).

Schuetze, Hinrich and Jan Pedersen (1995). ***Information Retrieval Based on Word Senses*** (cit. on p. 11).

Schulman, Ari (2009). ‘Why Minds Are Not Like Computers’. In: ***The New Atlantis*** 23, pp. 46–68.

Searle, John (1980). ‘Minds, Brains, and Programs’. In: ***Behavioral and Brain Sciences*** 3.3, pp. 417–457.

Shattuck, Roger (1959). ***The Banquet Years***. London: Faber.

Shu, Bo and Subhash Kak (1999). ‘A neural network-based intelligent metasearch engine’. In: ***Information Sciences*** 120 (cit. on p. 15).

Singh, Push (2005). ‘EM-ONE: An Architecture for Reflective Commonsense Thinking’. PhD thesis. Massachusetts Institute of Technology.

Srinivasan, P (2001). ‘Vocabulary mining for information retrieval: rough sets and fuzzy sets’. In: ***Information Processing and Management*** 37.1, pp. 15–38 (cit. on p. 11).

Stahl, Bernd Carsten, Marina Jirotka and Grace Eden (2013). ‘Responsible Research and Innovation in Information and Communication Technology: Identifying and Engaging with the Ethical Implications of ICTs’. In: ***Responsible In-***

novation. Ed. by Richard Owen. John Wiley and Sons. Chap. 11, pp. 199–218.

Sternberg, Robert J (1999). ***Handbook of creativity***. Cambridge University Press, p. 490.

– (2006). 'The Nature of Creativity'. In: ***Creativity Research journal*** 18.1, pp. 87–98.

Sutcliffe, Alistrair and Mark Ennis (1998). 'Towards a cognitive theory of information retrieval'. In: ***Interacting with Computers*** 10, pp. 321–351 (cit. on p. 11).

Taye, Mohammad Mustafa (2009). 'Ontology Alignment Mechanisms for Improving Web-based Searching'. PhD thesis. De Montort University (cit. on p. 15).

Thomas, Sue et al. (2007). 'Transliteracy: Crossing divides'. In: ***First Monday*** 12.12.

Turing, Alan (1950). 'Computing Machinery and Intelligence'. In: ***Mind*** 59, pp. 433–460.

Varshney, Lav R et al. (2013). 'Cognition as a Part of Computational Creativity'. In: ***12th International IEEE Conference on Cognitive Informatics and Cognitive Computing***. New York City, USA, pp. 36–43.

Ventura, Dan (2008). 'A Reductio Ad Absurdum Experiment in Sufficiency for Evaluating (Computational) Creative Systems'. In: ***5th International Joint Workshop on Computational Creativty***. Madrid, Spain.

Vian, Boris (2006). ***'Pataphysics? What's That?*** Trans. by Stanley Chapman. London: Atlas Press.

Vries, Erica de (1993). 'Browsing vs Searching'. In: ***OCTO report 93/02*** (cit. on p. 12).

Walker, Richard (2012). ***The Human Brain Project***. Tech. rep. HBP-PS Consortium.

Wallas, Graham (1926). ***The Art of Thought***. Jonathan Cape.

Walsh, Dave (2001). ***Absinthe, Bicycles and Merdre***.

Wickson, F., A.L. Carew and A.W. Russell (2006). 'Transdisciplinary research: characteristics, quandaries and quality'. In: ***Futures*** 38.9, pp. 1046–1059 (cit. on p. 33).

Widyantoro, D.H. and J. Yen (2001). 'A fuzzy ontology-based abstract search engine and its user studies'. In: ***10th IEEE International Conference on Fuzzy Systems*** 2, pp. 1291–1294 (cit. on pp. 11, 15).

Wiggins, Geraint A (2006). 'A preliminary framework for description, analysis and comparison of creative systems'. In: ***Knowledge Based Systems*** 19.7, pp. 449–458.

Yang, Hongji (2013). 'Editorial'. In: ***International journal of Creative Computing*** 1.1, pp. 1–3.

Yossarian (2015). ***Yossarian***.

KTHXBYE

Go to TOC