List of Todos

Institute of Creative Technologies
De Montfort University

Fania Raczinski

Web Collider

A pataphysical methodology for applying creativity to exploratory search

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: 21st June 2015

PRE☺

Abstract

Web Collider - Fania Raczinski

A pataphysical methodology for applying creativity to exploratory search

Creativity, Pataphysics and Computers

Obscure French Philosophy

This paper looks at defining, analysing and practicing how creativity can be applied to search tools. It defines creativity with respect to search and discusses how these concepts could be applied in software engineering using principles from the pseudo-philosophy of pataphysics. The aim of the proposed tool is to generate surprising, novel, humorous and provocative search results instead of purely relevant ones, in order to inspire a more creative interaction between a user, their information need and the application. A proof-of-concept prototype is described to justify the ideas presented before issues and future work are discussed.

Contents

List of Figures

List of Tables

List of Equations

List of Code Snippets

Acronyms

AI Artificial Intelligence. 51, 54

CC Creative Computing. 43, 54, 55, 56, 57

DH Digital Humanities. 55, 57, 60

ICCC International Conference on Computational Creativity. 54

IJCrC International Journal of Creative Computing. 55

IN Information Need. 68, 69

IR Information Retrieval. 71, 72, 74, 80

NLP Natural Language Processing. 82

NLTK Natural Language Toolkit. 81, 125

Part I

HΣLLΘ WΘRLD

Chapter 1

Introduction

§¶ *Part of this chapter has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.*

HELLO WORLD

This thesis describes a general study into computer creativity and in particular a pataphysical methodology applied to creative information retrieval.

A pataphysical methodology for applying creativity to exploratory search.

A semantic[1] search[2] tool based on a pataphysical[3] algorithm[4] using a 'patadata'[5] ontology[6]

1. Semantic Web technologies
2. IR techniques
3. Pataphysics
4. Ranking algorithms
5. Data - metdadata - patadata
6. Ontologies, RDF, OWL, etc

finish introduction

In the following chapters, I will examine how pataphysics and creativity map onto one another, give an outline of the field of information retrieval...

1.1 Problem / Motivation / Context

Assumptions: computers can be creative.

complete assumptions

From relevant to creative.

Why pataphysics? Pataphysics is highly subjective and particular and is as such very suitable for this kind of transformation from relevant to creative. Pataphysics can provide some useful techniques that are very suitable for creative computing.

> Purposive without purpose (Kant)

The aim of this project, in simple words, is to design a tool for creative searching on the Web. It tries to create a fresh way of searching and navigating through information and content on the Internet, to bring back some of the inspirational chance encounters that were so characteristic for libraries.

Current information retrieval systems might be used for creative purposes, however, they do not directly provide creative results to their users, instead they focus on precise and relevant results only. Therefore we argue that a new style of system is required. It is clear that the fundamental problem in this is that standard algorithms are not suited for these problems, with them considering a document to be groupings of words in traditional IR systems, and that an entire document falls under the same classifications in semantic IR systems.

The tool will challenge the way we understand data and metadata and suggest new meanings between them in accordance with the underlying philosophy. The way we think about information is redefined. It will create new standards or rather get rid of all previous standards and classifications. It will create new links and connections between pieces of information that are not simply based on keyword similarities but rather a more poetic sense of unity. It will be a refreshing new view compared to the structured and standardised thinking of computer scientists. It will challenge the binary logic that dominates the world of programmers and concentrate on what they would view as an illogical and unstructured system going against all standards.

Jorge Luis Borges 'Chinese Encyclopaedia' (Borges 2000) illustrates this idea very well. The encyclopaedia lists the following results under the category of 'animal'.

1. those that belong to the Emperor,
2. embalmed ones,
3. those that are trained,
4. suckling pigs,
5. mermaids,
6. fabulous ones,
7. stray dogs,
8. those included in the present classification,
9. those that tremble as if they were mad,
10. innumerable ones,
11. those drawn with a very fine camelhair brush,
12. others,
13. those that have just broken a flower vase,
14. those that from a long way off look like flies.

Although these are all perfectly valid results, it is clear that they form a more creative, even poetic, view of what an animal might be than the Oxford English Dictionary's prosaic: "a living organism which feeds on organic matter".

We have introduced the motivation and concept for a creative Web search tool and discussed some of the major challenges a project like this faces. With Web search being a major research and learning tool nowadays, it is imperative to think about how such a tool could be (ab)used. Ethical issues that arise through the provision of unexpected results, and the misunderstandings this could lead to, will be discussed in future work. Nevertheless, we believe that creative Web search can facilitate inspirational learning through an exploratory search journey and we hope our tool will provide just that.

1.1.1 Related Work

The Syzygy Surfer

The research presented here is built on the initial ideas of Jim Hendler and Andrew Hugill's "Syzygy Surfer" (Hendler and Hugill 2011, 2013). They first suggested the use of three pataphysical principles, namely clinamen, syzygy and anomaly, to create a new type of Web search engine, reminiscent of the experience of "surfing the Web". This is in contrast to current search engines which value relevant results over creative ones. "Surfing" used to be a creative interaction between a user and the web of information on the Internet, but the regular use of modern search engines has changed our expectations of this sort of knowledge acquisition. It has drifted away from a learning process by exploring the Web to a straightforward process of information retrieval similar to looking up a word in a dictionary.

> The ambiguity of experience is the hallmark of creativity, that is captured in the essence of pataphysics. Traversing the representations of this ambiguity using algorithms inspired by the syzygy, clinamen and anomaly of pataphysics, using a panalogical mechanism applied to metadata, should be able to humanize and even poeticize the experience of searching the Web.(Hendler and Hugill 2013)

Their inspirations come from Borges' "Chinese Encyclopaedia" (Borges 2000) (for the underlying poetic sense of unity), Jarry's Pataphysics (Jarry 1996) (for the concept of patadata – data beyond metadata) and panalogies (parallel analogies – to introduce ambiguity, since it allows various descriptions of the same object) as formulated by Singh (Singh 2005).

Yossarianlives

> Traditional search like Google, Bing, Yahoo, or DuckDuckGo returns the most popular results, and gives you expected & cliché ideas. It's what everyone else already thinks about a topic.
>
> Yossarian creative search returns diverse and unexpected results that share loose associations to your search. The results help you think about your topic in new ways and generate new ideas.
>
> If you search "beauty" in Google you get pages and pages of white models. It returns a singular way of thinking about the topic.
>
> Search "beauty" in Yossarian it returns disparate results with shared attributes, showing many different ways of thinking. Is beauty a diamond (strong, rare, flawless), or is family beautiful (activity, togetherness) or is beauty architected (designed, planned, precision)?[1]
>
> Use traditional search for when you know what you are looking for.
>
> Traditional search is for learning what the world already knows.

[1] https://yossarianlives.com/

Use creative search for when you don't know what you are looking for.

Creative search is for helping you come up with new ideas.[2]

Augmented creativity

Seeing lateral connections is an uniquely human ability, and your individual ability to make these connections is based on your own experience. We build tools that help you see more by suggesting connections that you probably wouldn't have considered otherwise, increasing the diversity and frequency of your ideas, and priming you for creativity.

Lateral Discovery

When working to escape filter bubbles too often randomenss and complete seridipity are seen as the answer for new discovery. We build tools that explore lateral association between content and believe this approach can lead to predictable moments of discovery that drive engagement, traffic, and purchase, etc.

Metaphorical Search

Search engines today have a problem in that they tell us what the world already knows, reinforcing existing knowledge. We build tools that seek to remedy this by returning results that are disperate, but metaphorically related to the query. These types of results are incredably useful for any one who derives value from new ideas.

Creative Graph

Facebook created the Social Graph helping us understand how people and things are connected, Google created the Knowledge Graph to codify the worlds information, and here at Yossarian Lives we are developing the Creative Graph, a radical new way to understand the conceptual relationships between things.[3]

The Library of Babel

The Library of Babel is a place for scholars to do research, for artists and writers to seek inspiration, for anyone with curiosity or a sense of humor to reflect on the weirdness of existence - in short, it's just like any other library. If completed, it would contain every possible combination of 1,312,000 characters, including lower case letters, space, comma, and period. Thus, it would contain every book that ever has been written, and every book that ever could be - including every play, every song, every scientific paper, every legal decision, every constitution, every piece of scripture, and so

[2] https://yossarianlives.com/

[3] http://about.yossarianlives.com/index.html

on. At present it contains all possible pages of 3200 characters, about 104677 books.

Since I imagine the question will present itself in some visitors' minds (a certain amount of distrust of the virtual is inevitable) I'll head off any doubts: any text you find in any location of the library will be in the same place in perpetuity. We do not simply generate and store books as they are requested - in fact, the storage demands would make that impossible. Every possible permutation of letters is accessible at this very moment in one of the library's books, only awaiting its discovery. We encourage those who find strange concatenations among the variations of letters to write about their discoveries in the forum, so future generations may benefit from their research.[4]

1.1.2 Research Questions

- How can we make a search tool that is inspirational rather than informational?
- How can we get search results that are unexpected and yet make sense?
- How can we rank search results but still be true to Pataphysics philosophy?
- How can we represent and structure data to reflect its context, meaning and subjectivity?
- How can we present search results in a creative and pataphysical way?
- How does Pataphysics relate to creative computing?
- How can we use Pataphysics as inspiration for search ranking?
- How can we write a specifically creative algorithm?
- How can Semantic Web technologies help with the representation of patadata?
- What does it mean for search results to be creative/relevant?
- Can computers be creative?
- What does it actually mean to be creative even for a human being, etc
- Is pataphysics creative?
- What is a relevant search result?
- Is creativity irrelevant?

answer research questions in conclusion

[4] https://libraryofbabel.info/

1.2 Methodology

This project combines research in science and art. It is an interdisciplinary research project.

This project has roots in disciplines such as Computer Science and Humanities.

Information Retrieval
: Software Engineering, Semantic Web

Pataphysics
: Literature, Philosophy, Ontology

Creativity
: Cognitive Science, Artificial Intelligence

In regards to my project:

- A concept implementation method is used with a descriptive-other approach
- A qualitative investigation into if and why the proposed search results are useful will be done
- Following experimental methodologies, to evaluate the proposed new solution to the problem of creative search

Epistemology
: Subjective/Argumentative

Methodology
: Experimental, Interpretative, Qualitative

Methods
: Concept implementation, (Heuristic) Evaluation

1.3 Deliverables / Outcomes

- Design a tool for creative searching on the Web
- Design pataphysics inspired algorithms to model creativity in this tool
- Produce a proof-of-concept prototype
- Propose a framework for evaluating and interpreting creative search results

!!! http://pata.fania.eu

1.3.1 Contribution to Knowledge

Compared to the mainstream search engines available publicly, my project's approach to searching will be very different. Google, for example, ranks web pages based on the number and quality of incoming hyperlinks (Google 2012), while I will be using the concept of patadata (data-metadata-patadata) and semantic web technologies.

There have been interdisciplinary projects like Johanna Drucker's Speclab (Drucker 2009) which have similar inspirations, and literature on Pataphysics is found in several places (Bök 2002; Hugill 2012a) although the idea of using it as an inspiration for a search engine appears to be very new (Hendler and Hugill 2013).

A lot of literature exists on how to model creativity on a computer [5] [7], explaining several theories on how the mind works and how to simulate this on a machine. However none of these have any kind of pataphysical inspiration and aren't applicable to search engines, so we have presented a paper on creativity in search results just recently (Raczinski et al. 2013).

1.3.2 Publications

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. (Sawle et al. 2011)

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. (Hugill et al. 2013)

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. (Raczinski et al. 2013)

1.4 The Hitchhiker's Guide to this Thesis

PREFACE

Part I

IN THE BEGINNING...

Chapter 1
: Introduction

Chapter 2
: Methodology

Part II
: IN A GALAXY FAR FAR AWAY...

Chapter 3
: Pataphysics

Chapter 4
: Creativity and Computers

Chapter 5
: IR and NLP

Part III
: THE CORE: TECHNO-LOGIC

Chapter 6
: Theoretical Foundations

Chapter 7
: Practical Implementation

Chapter 8
: Impact and Applications - Case Study

Part IV
: INTECHNOIL-LOGICALYSIS

Chapter 9
: Interpretation / Evaluation

Chapter 10
: Patacritical Analysis

Part V
: HAPPY END

Chapter 11
: Aspirations

Chapter 12
: Observations

POSTFACE

Chapter 2

Methodology

Description and justification of methodology. . .

> "Only those who attempt the absurd achieve the impossible." (attributed to M.C. Escher)

Ernest Edmonds, Linda Candy, 2010. "Relating Theory, Practice and Evaluation in Practitioner Research" LEONARDO, Vol 43, No 5, pp470-476.

Epistemology
: "A broad and high-level outline of the reasoning process by which a school of thought performs its empirical and logical work." Wikipedia

Methodology
: "Less high level than epistemology is methodology. It refers to a more specific manner in which to do empirical and logical work. The same epistemology can have several methodologies." Wikipedia

Method
: A methodology can consist of several methods. Wikipedia

Epistemology

"is the branch of philosophy concerned with the nature and scope (limitations) of knowledge. It addresses mainly the following questions: What is knowledge? How is knowledge acquired? To what extent is it possible for a given subject or entity to be known?" [2]

"A broad and high-level outline of the reasoning process by which a school of thought performs its empirical and logical work." [1]

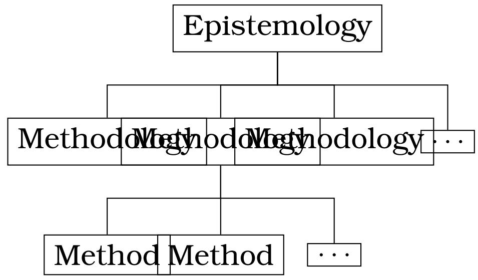

Figure 2.1: Epistemology breakdown chart

Methodology

"is usually a guideline system for solving a problem, with specific components such as phases, tasks, methods, techniques and tools. It can be defined also as follows: 1. 'the analysis of the principles of methods, rules, and postulates employed by a discipline' 2. 'the systematic study of methods that are, can be, or have been applied within a discipline' 3. 'the study or description of methods'" [2]

"Less high level than epistemology is methodology. It refers to a more specific manner in which to do empirical and logical work. The same epistemology can have several methodologies." [1]

Research Strategy is a procedure for achieving a particular intermediary research objective—such as sampling, data collection, or data analysis. We may therefore speak of sampling strategies or data analysis strategies. [2]

Research Approach refers to an integrated set of research principles and general procedural guidelines. Examples of research approaches include experiments, surveys, correlational studies, ethnographic research, and phenomenological inquiry. [2]

Qualitative researchers aim to gather an in-depth understanding of human behavior and the reasons that govern such behavior. The qualitative method investigates the why and how of decision making, not just what, where, when. Hence, smaller but focused samples are more often needed than large samples. [2]

Quantitative research refers to the systematic empirical investigation of social phenomena via statistical, mathematical or computational techniques. The objective of quantitative research is to develop and employ mathematical models, theories and/or hypotheses pertaining to phenomena. Quantitative data is any data that is in numerical form such as statistics, percentages, etc. [2]

2.1 Intradisciplinary Research

Different disciplines prefer different research methodologies. It makes sense that research in medicine, chemistry, literature or mathematics all use different methods. What could a mathematician achieve in a white laboratory coat and test tubes in his hand, and similarly, what could a chemist achieve with pen, paper and a calculator?

> "methodological pluralism is acceptable but what is not acceptable is philosophical pluralism" ??

What would be traditional RM in those fields?
Why can I not mix and match them?
What do I do now/instead?
Can inter/multi/trans-disciplinary research be NOT collaborative but done by a single person?

> "When you describe your methods it is necessary to state how you have addressed the research questions and/or hypotheses. The methods should be described in enough detail for the study to be replicated, or at least repeated in a similar way in another situation. Every stage should be explained and justified with clear reasons for the choice of your particular methods and materials." [1]

2.1.1 Computer Science Research

In their rather old but still insightful analysis of over 600 papers, published between 1995 and 1999, Ramesh et al (Ramesh et al. 2004) have shown that -by far- the most common approach to research in computer science was "formulative" (as opposed to "descriptive" and "evaluative") in particular in regards to "processes, methods and algorithms".

```
Research Approach in CS [4]
Descriptive: (9.88%)
```
- `Descriptive-system (4.14%)`
- `Descriptive-other (5.10%)`
- `Review of literature (0.64%)`

[1] http://bit.ly/1Edj84y

Evaluative: (10.98%)

- Evaluative-deductive (1.11%)
- Evaluative-interpretive (-)
- Evaluative-critical (-)
- Evaluative-other (9.87%)

Formulative: (79.15%)

- Formulative-framework (2.39%)
- Formulative-guidelines/standards (0.64%)
- Formulative-model (5.73%)
- Formulative-process, method, algorithm (52.55%)
- Formulative-classification/taxonomy (0.80%)
- Formulative-concept (17.04%)

Research Method in CS [4]

- Action research (-)
- Conceptual analysis (15.13%)
- Conceptual analysis/mathematical (73.41%)
- Concept implementation (2.87%)
- Case study (0.16%)
- Data analysis (0.16%)
- Discourse analysis (-)
- Ethnography (-)
- Field experiment (-)
- Field study (0.16%)
- Grounded theory (-)
- Hermeneutics (-)
- Instrument development (-)
- Laboratory experiment (human subjects) (1.75%)
- Literature review / analysis (0.32%)
- Laboratory experiment (software) (1.91%)
- Meta-analysis (-)
- Mathematical proof (2.39%)
- Protocol analysis (-)
- Phenomenology (-)
- Simulation (1.75%)
- Descriptive/exploratory survey (-)

Formal

Formal methodologies are mostly used to prove facts about algorithms and system. Formal specification of a software component in order to allow the automatic verification of an implementation of that component, the time or space complexity of an algorithm, or on the correctness or the quality of the solutions generated by the algorithm. [5]

Experimental

Experimental methodologies are broadly used in CS to evaluate new solutions for problems. Experimental evaluation is often divided into two phases. In an exploratory phase the researcher is taking measurements that will help identify what are the questions that should be asked about the system under evaluation. Then an evaluation phase will attempt to answer these questions. A well-designed experiment will start with a list of the questions that the experiment is expected to answer. [5]

Build

A "build" research methodology consists of building an artifact — either a physical artifact or a software system — to demonstrate that it is possible. To be considered research, the construction of the artifact must be new or it must include new features that have not been demonstrated before in other artifacts. [5]

Process

A process methodology is used to understand the processes used to accomplish tasks in Computing Science. This methodology is mostly used in the areas of Software Engineering and Man-Machine Interface which deal with the way humans build and use computer systems. The study of processes may also be used to understand cognition in the field of Artificial Intelligence. [5]

Model

The model methodology is centered on defining an abstract model for a real system. This model will be much less complex than the system that it models, and therefore will allow the researcher to better understand the system and to use the model to perform experiments that could not be performed in the system itself because of cost or accessibility. The model methodology is often used in combination with the other four methodologies. Experiments based on a model are called simulations. When a formal description of the model is created to verify the functionality or correctness of a system, the task is called model checking. [5]

[14]: Four quadrant model: 1. What do we want to achieve? 2. Where does the data come from? 3. What do we do with the data? 4. Have we achieved our goal? Iterative process, can repeat etc.

2.1.2 Humanities Research

2.1.3 Arts Research

2.2 Transdisciplinary Research

Multidisciplinarity
: "concerns itself with studying a research topic in not just one discipline but in several simultaneously."

Interdisciplinarity
: "has a different goal than multidisciplinarity. It concerns the transfer of methods from one discipline to another."

Transdisciplinarity
: "concerns that which is at once between the disciplines, across the different disciplines, and beyond all disciplines."

(Nicolescu 2010)

Problem Focus: (solve complex, multi-dimensional, particular problems)

"TD research therefore starts with a problem that is 'in the world and actual' as opposed to 'in my head and conceptual'." "This inherent feature of 'creating change' highlights the relevance of using the term 'consequential' to characterise TD research approaches and problems. " [9]

2.2.1 Practice Based Research

2.3 Research Approach

Part II

IN A G∀L∀XY F∀R F∀R AW∀Y . . .

Chapter 3

’Pataphysics

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

> To understand ’pataphysics is to fail to understand ’pataphysics. (Hugill 2012a)

[1]

It is probably impossible to define pataphysics in one sentence. There is no definition that does justice to what pataphysics really means and no single definition that is truer than any other. In fact, the college of pataphysics in France itself has published a book (Brotchie et al. 2003) with over 100 definitions that they all call equally valid. This chapter therefore begins with a selection of definitions from that book (quoting their original sources).

> Pataphysics . . . is the science of that which is superinduced upon metaphysics, whether within or beyond the latter’s limitations, extending as far beyond metaphysics as the latter extends beyond physics. [. . .] Pataphysics will be, above all, the science of the particular, despite the common opinion that the only science is that

[1] Although note how the perplexing apostrophe that sometimes appears before the word ‘pataphysics undermines too literal an interpretation of this construction. Jarry only ever used the apostrophe on a single occasion, specifying that he did so “in order to avoid a simple pun”. What that pun might be has never been fully explained.

of the general. Pataphysics will examine the laws governing exceptions, and will explain the universe supplementary to this one. [. . .] DEFINITION: Pataphysics is the science of imaginary solutions, which symbolically attributes the properties of objects, described by their virtuality, to their lineaments." (Alfred Jarry, "Exploits and Opinions of Dr Faustroll, Pataphysician" written in 1897-8 and published posthumously in 1911) (Jarry 1996)

'Pataphysics is patient; 'Pataphysics is benign; 'Pataphysics envies nothing, is never distracted, never puffed up, it has neither aspirations nor seeks not its own, it is even-tempered, and thinks not evil; it mocks not iniquity: it is enraptured with scientific truth; it supports everything, believes everything, has faith in everything and upholds everything that is. ("Épanorthose sur le Clinamen moral", Cahiers du Collège de 'Pataphysique, 21, 22 Sable 83 (29 December 1955 vulg.)) (Brotchie et al. 2003)

'Pataphysics passes easily from one state of apparent definition to another. Thus it can present itself under the aspect of a gas, a liquid or a solid. (Patafluens 2001, Istituto Patafisico Vitellianese, Viadana, 2002) (Brotchie et al. 2003)

'Pataphysics, 'the science of the particular', does not, therefore, study the rules governing the general recurrence of a periodic incident (the expected case) so much as study the games governing the special occurrence of a sporadic accident (the excepted case). [. . .] Jarry performs humorously on behalf of literature what Nietzsche performs seriously on behalf of philosophy. Both thinkers in effect attempt to dream up a 'gay science', whose joie de vivre thrives wherever the tyranny of truth has increased our esteem for the lie and wherever the tyranny of reason has increased our esteem for the mad. (Christian Bök, 'Pataphysics, The Poetics of an Imaginary Science, Northwestern University Press, 1Il., 2002) (Bök 2002)

La pataphysique est la fin des fins.	'Pataphysics is the end of ends.
La pataphysique est la fin des faims.	'Pataphysics is the end of hunger.
La pataphysique est la faim des fins.	'Pataphysics is the hunger for ends.
La pataphysique est le fin du fin.	'Pataphysics is the finest of the fine.

> (The first motto is that of the official Collège notepaper. Its three variants have appeared in elsewhere in Collège publications. - Collège de 'Pataphysique) (Brotchie et al. 2003)

> The branch of philosophy that deals with an imaginary realm additional to metaphysics. (Oxford Dictionary)

> Epiphany – "to express the bursting forth or the revelation of pataphysics" Dr Sandomir (Hugill 2012a, p.174)

I divided my research about pataphysics into four approaches. The first: learn about its inventor - Alfred Jarry. The second: read his work. The third: read what others have to say about pataphysics. The fourth: read other literature that could be classed as pataphysical. Eventually I ended up seeing pataphysics in everything, there was no escaping it anymore. I had turned a Pataphysician.

Personally, when I try to explain pataphysics to laymen, I use the last scene of the movie 'Men In Black' as an example. The scene zooms out further and further, from a close up of Will Smith, to his car, to a shot of the city from above, to a shot of the earth, the galaxy, the universe and finally it is revealed that the universe is in a marble that is being toyed with by an alien. This is a good example of different layers of abstraction – Will Smith in his car represents the physical layer, the universe the metaphysical layer and the alien marble the pataphysical layer. The outro scene can be seen on YouTube [2].

3.1 Conscious

> Jarry was "attempting to transcend his own existence." (Hugill 2012b)

> It is certainly true that making life 'as beautiful as literature' was one of his goals. (Hugill 2012b)

Alfred Jarry was born in Laval, Mayenne, France in 1873 and died in Paris in 1907, at the age of 34. He was known as a poet, dramatist, novelist and journalist but also as a graphic artist. His hobbies included entomology, fishing, cycling, fencing, shooting and drinking.

[2] http://youtu.be/1QPl1-TKaEE

Figure 3.1: Men in Black Screenshots of Ending Sequence

Figure 3.2: FelixVallotton + jarry + jpicasso

He went to school in Rennes, where his physics teachers Félix-Frédéric Hébert left such a big impression on Jarry that he would later be his inspiration for Père Ubu. He passed his baccalauréat with 17 and moved to Paris to attend the lycée Henri IV in preparation to apply for admission to the École Normale Supérieure but eventually gave upon the entrance exam after several unsuccessful attempts. He met another teacher at the lycée, this time a philosophy teacher called Henri Bergson, who inspired him greatly. He published his first collection of poems in 1893, aged 20, the year his mother died. One of his classmates there described him as follows.

> [...] I found Jarry's mental processes disturbing. When he let himself go he seemed in thrall to a torrent of words outside his con-

> trol. It was no longer a person speaking, but a machine controlled by a demon. His staccato voice, metallic and nasal, his abrupt puppet-like gestures, his fixed expression and uncontrolled flood of language, his grotesque and brilliant turns of phrases, ended up provoking a feeling of disquiet. He was informed, intelligent, and discriminating; he was good person, secretly kind, perhaps even shy beneath it all [. . .] but his originality resembled nothing short of a mental anomaly. (Jarry's classmate at the lycée Henri IV: Gandilhon Gens-d'Armes "Alfred Jarry au lycée Henri IV" Les Marges, XXIII, 91 (15 Jan 1922) as quoted in (Brotchie 2011))

He was at the centre of the avant-garde movement in Paris around that time, at the centre of the Tuesday meetings of the Mercure de France (a literary magazine run by Alfred Valette and his wife Rachilde, who soon became a sort of substitute family to Jarry who was roughly 15 years younger than them). Being rather misogynist at times and homosexually inclined, Rachilde was one of his very few female friends.

The following year, 1895, he briefly joined the army in the 101st Infantry, after having dodged it by being an enrolled student at the lycée. He followed rules there pedantically but hated the loss of his individualism. According to Brotchie, he "chose subservience, but subservience taken to the point of parody: the pataphysical solution to the problem of obedience" (Brotchie 2011). Probably the only thing he enjoyed there was the fencing and shooting training. He looked funny in the uniform that was too big for him being so small (5'3") so he was eventually excused from parades and after a few months he was allowed to leave to Paris frequently. He was discharged in December 1895 on medical grounds: gallstones. It is not unlikely that he faked the illness by drinking picric acid.

His father had died just two months earlier and had left him a small inheritance, which he spent mostly on publishing his very own magazine dedicated to symbolist wood carvings, the Perhinderion. He had previously co-edited the magazine L'Ymagier with Remy de Gourmont between 1893 and 1894. He joined Aurélien Lugné-Poë as his secretary (his only ever real job) at the Théâtre de l'Œuvre after his discharge at the army, where he would pour his utmost attention to putting his Ubu play on the stage. He also played a small role in the production of Peer Gynt at the Œuvre earlier in 1896. The printed version of Ubu Roi appeared in Le Livre d'Art in the middle of the year with Jarry's carved woodcut image of Ubu that became so popular. The première took place on 10th December that year and caused an outrage in the audience after the first word: "merdre" (sometimes translated as "pshit"). Jarry had previously arranged for certain friends to counter any reaction of the general audience and to prevent under all circumstances for the play to reach its conclusion. The performance

Figure 3.3: Woodprint of Ubu by Alfred Jarry

went according to plan. The uproar after the first word was uttered was immense, the performance had to be interrupted at times to calm the audience and it finished in shouts of praise, protest and insults. There were no further performances but the event was considered historic even at the time and is now widely seen as the first "modern" play (Brotchie 2011, p.168-169). And as Dave Walsh puts it: "Movements such as Dadaism, Surrealism, Futurism, Expressionism Cubism, Theatre of the Absurd - all owe debts to [Jarry's] works." (Walsh 2001)

Although Ubu's mannerism of speech was originally imitating Jarry's, as suggested by Lugné-Poë (Brotchie 2011, p.155), Jarry continued to adapt Ubu's mannerisms.

> "Those who knew him said that his nauseating appearance hid a youth who was stubborn yet shy, proud and little full of himself, but good-natured and ingenuous behind his cynicism, one who was fiercely independent and rigorously honest." (Henri de Régnier, as quoted in (Brotchie 2011, p.181))

> "Alfred Jarry had a very particular way of speaking to that was disconcerting to those who heard it for the first time. He said 'we', when referring to himself, and substituted verbs for nouns, in imitation of ancient Greek. Example: 'celui qui soufflé' (that which blows) for the wind, and 'celui qui se traîne' (that which crawls along) for the train, even if it was an express! This made conversation somewhat complicated, not least because of the rapidity of

> his delivery." (Rachilde, as quoted in (Brotchie 2011, p.181))

> "Alfred Jarry was a man of letters to an unprecedented extent. His smallest actions, his childish pranks, everything he did was literature. His whole life was shaped by literature, and only by literature." (Apollinaire, as quoted in (Brotchie 2011, p.307))

Jarry spent the next few years writing. He had spent all his inheritance on the publication of his magazine and the production of Ubu Roi. It is during this time that he moved to his infamous tiny flat on the second-and-a-half floor. Jarry could just about stand upright but any guests had to crouch. He had no electricity or gas and no means of cooking (Brotchie 2011, p.195). In December 1897 he formed a marionette theatre with his friend Claude Terasse: the Théâtre de Pantins and they performed Ubu Roi in January 1898 without riots in the audience.

Jarry then gradually withdrew from the literary circles in Paris and spent more time in a little shack on the banks of the Seine near the village of Le Coudray. He started writing a regular review column for the Revue Blanche in 1900, the income of which he certainly needed much. There was a brief revival of the Ubu marionette play in the Cabaret des Quat'z'Arts in 1901.

Around 1904 he began drinking ether, the absinthe not strong enough anymore. In the winter of 1905 he was very ill, the cold and poverty not helping. In 1906, his friends became more and more concerned about his deteriorating health and eventually Valette and Saltas sent him to his sister Charlotte. He then spent some time in Paris and some in Laval at his sister's place over the next year. Jarry then died in November of 1907 of meningeal tuberculosis. His last request was for a toothpick.

> "He believes that the decomposing brain goes on working after death and it is its dreams that are Paradise." (Jarry 1906 in a letter to Rachilde (Brotchie and Chapman 2007) - 'he' refers to himself)

Studying Jarry's life gives certain insights into the man who created pataphysics and why he might have done so. Alastair Brotchie has written the probably most concise and recent biography of Alfred Jarry in English language and most of the information summarised here comes from this book (Brotchie 2011). Roger Shattuck gives a very nice introduction about Jarry in relation to the time and place he lived in, in his book "The Banquet Years" (Shattuck 1959). However, he does not focus on Jarry alone but rather on the time period and four personalities (Alfred Jarry, Henri Rousseau, Erik Satie and Guillaume Apollinaire) he chose as representatives of the era.

His Writing

Jarry has written a good amount of texts in his short life and he didn't confine himself to a single category either. He wrote poems, novels, short stories, essays, art reviews, theatre reviews and plays and also produced translations of a few texts into French. Many of his texts were completely fictional, some had autobiographical aspects and some scientific and most of them had a sarcastic sense of humour. Trying to summarise Jarry's style of writing or attempting to interpret his whole body of work seems impossible though.

> "Jarry was an acknowledged classical scholar, had already worked as a reviewer of art and drama, had edited two art magazines, was up to date with modern scientific theory, especially physics, read widely in mathematics and psychology, and had an extensive basic knowledge of philosophy." (Brotchie 2011)

James A Cutshall says that "instead of Jarry the man and the meaning of his literary endeavours becoming clearer with the passage of time, both have become increasingly indistinct" (Cutshall 1988, p.246). The intention of his thesis was to show the seriousness implied behind the humour in many of Jarry's novels, in order to give the author the merit he deserved. Cutshall wrote about Jarry's novels rather than simply seeing him as the playwright of the Ubu plays. He surveyed existing criticism about Jarry's texts and provided his own view on them. He immortalised Jarry by saying "whether or not this is the sort of 'éternité' sought by the heroes of Jarry's novels, it is certainly that which their author somewhat belatedly has found" (Cutshall 1988, p.248).

3

Cutshall was not the only one who has written about certain less-known texts by Jarry. Marieke Dubbelboer's thesis "Ubusing Culture" is also interesting in this regard since it concentrates completely on the "Almanachs du Père Ubu" (published in 1898 and 1901) (Dubbelboer 2009). She was looking for keys to Jarry's poetics in those texts, which she says "seemed to defy labelling or literary

[3] http://bit.ly/1QOVZW9

Figure 3.4: Faustroll illustration by Steve Morrison

norms" (Dubbelboer 2009, p.10). She claims the Almanacs to be quite radical and exemplary of his innovative poetics moving away from Symbolism and towards the Avant-Garde. In general she says his work "can be characterized as playful, elusive, paradoxical and provocative" (Dubbelboer 2009, p.197) and his two Almanacs are the essence of his non-conformist attitude. They were written at a time of change for Jarry, when he withdrew from his usual circles in Paris and he published in new magazines, which links his change in writing according to Dubbelboer.

A list of his works can be found in the appendix **??**.

3.2 Self-conscious

We will need to understand the essence of pataphysics to understand how it relates to creative computing.

Jarry first defined pataphysics in his book "Exploits and Opinions of Dr Faustroll, Pataphysician" written in 1898 and published posthumously in 1911 (Jarry 1996). But the concept appeared as early as in 1893 in his prose text Guignol that won him a prize in the newspaper L'Echo de Paris and it appears in many of his writings. He originally intended to write a whole book called "Elements of Pataphysics" but only part of this appeared in Faustroll.

Zoe Corbyn gives a very simple short introduction for beginners of the topic in an article in the Guardian (Corbyn 2005) in 2005. She describes it like this:

> Correct definitions are equivalent to wrong ones; all religions are on a par as imaginary and equally important; chalk really is cheese. It's an escape from reality - reminding us of just how idiotic the rules that dog our everyday existence are. (Jarry 1996)

Jean Baudrillard has a few other definitions for pataphysics in his text(Baudrillard 2007). According to him pataphysics is "the highest temptation of the spirit, the nail in the tire, the philosophy of the gaseous state, the science or the unique imaginary solution to the absence of problems" to name just a few.

Another rather strange interpretation of pataphysics is Asger Jorn's. He calls pataphysics a religion in the making (Jorn 1961). He claims that since "natural religion is the spiritual confirmation of material existence", "metaphysical religion represents the establishment of an ever deepening rift between material and spiritual life." He refers to the idea of equivalence in pataphysics and the absolute and links them to religion. He says "the great merit of Pataphysics is to have confirmed that there is no metaphysical justification for forcing everybody to believe in the same absurdity".

Cruickshank (Cruickshank nd) wrote a rather funny article on anti-matter. He links the creation of anti-matter atoms at CERN around 1996 with Jarry, saying that he had "beaten them to the punch" with his pataphysics.

Christian Bök (Bök 2002) tries to draw science and poetry together using pataphysics as the string that binds them. He compares Jarry and Nietzsche, saying Jarry performs humorously on behalf of literature what Nietzsche performs seriously on behalf of philosophy; both try to create an antiphilosophy (Bök 2002, p.9). He also claims that science and poetry have a similar history, undergoing the same four phases of distinct change but also that they have not evolved in sync with each other (Bök 2002, p.15).

Animalistic phase
: signs exist long before being known, they are written by nature

Mechanismic phase
: signs exist by being known, they are written by culture

Organismic phase
: signs evolve by being known, they are written across events by culture

Cyborganismic phase
: signs evolve beyond being known, they are written as events by culture

> Pataphysics is a surrational perspective that has had an extensive,

> yet forgotten, influence upon the canonic history of radical poetics. [...] Not only does this avant-garde pseudoscience valorise whatever is exceptional and paralogical; it also sets the parameters for the contemporary relationships between science and poetry. (Bök 2002, p.27)

Bök also compares Jarry and Nietzsche in regards to perspectivism (Bök 2002, p.31). For Nietzsche reality is the effect of a dream world in which "there are many kinds of truths, and consequently there is no truth". And similarly for Jarry, reality is an aspect of ethernity in which "there are only hallucinations, or perceptions," and every "perception is a hallucination which is true". Both argue that no view is absolute as well and pataphysics argues that every viewpoint is dissolute, including its own because no view can offer a norm. Even Jarry's ethernity is nowhere and somewhere at the same time.

In Faustroll, Bök says, "Jarry parodies the discourse of such scientific luminaries, who attempt to demonstrate the utility of science through the dramaturgic performance of a mechanical experiment" (Bök 2002, p.29).

According to the Collège de 'Pataphysique, it is convention to use the apostrophe at the beginning of the word ('Pataphysics) only in reference to Jarry's texts, to the science of imaginary solutions as such. Used as an adjective or in a more unconscious way it is written without the apostrophe. Jarry himself just indicated that the word is preceded by the apostrophe to avoid a pun.

- Vian, B. (2006). 'Pataphysics? What's That? (S. Chapman, Trans.). London: Atlas Press.(Vian 2006)
- Daumal, R. (2012). Pataphysical Essays. (T. Vosteen, Trans.). Cambridge, Massachusetts: Wakefield Press.(Daumal 2012)
- Brotchie, A. (Ed.). (1995). A True History of the College of 'Pataphysics - 1. (P. Edwards, Trans.). London: Atlas Press.(Brotchie 1995)

3.2.1 Symbology

Probably the most famous symbol of pataphysics is the grand gidouille, the big spiral on Ubu's fat belly. Not simply because it is a feature of Jarry's most popular creation but also because it represents one of the concepts of pataphysics itself: the antimony. The spiral can be interpreted as two spirals in one, the outer and the inner spiral. They represent the duality of pataphysics, the mutually incompatible in perfect harmony. The Collège de 'Pataphysique has adopted the spiral for its membership badges, in various colours and sizes for the different ranks of the college.

Another symbol of pataphysics is the green candle which refers to one of Jarry's last endeavours, published posthumously, a vast collection of his journalistic essays (Hugill 2012b). Some animals also symbolise pataphysics. The crocodile, the current vice-curator of the college is a crocodile named Lutembi (Hugill 2012b). Owls are another symbol; Jarry kept stuffed and live owls (Brotchie 2011, p.46)[13 p46] in his flat. The chameleon is another, having the ability to change colour and looking in two directions at the same time.

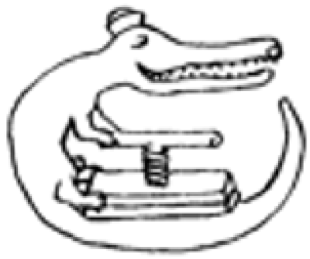

Figure 3.5: Crocodile from the CoP website

Figure 3.6: The Grand Gidouille

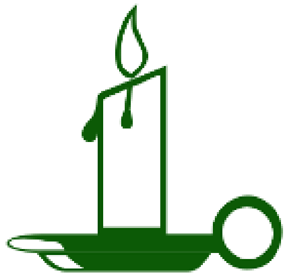

Figure 3.7: The green candle

We argue that pataphysics can facilitate creative computing. A pataphysical grammar consists of exceptions, syzygies, anomalies, clinamen, antinomies, contradictions, equivalents and imaginaries. Such constraints can transform the ways in which we may navigate and transform our conceptual space. Pataphysical concepts are likely to cause surprise and could therefore be considered unconventional and provocative.

3.2.2 Antimony

The antimony is the mutually incompatible. It appears everywhere in Jarry's writings. It represents the duality of things, the echo or symmetry, the good and the evil at the same time. Examples are the plus minus, the faust-troll, the haldern-ablou, the yes-but, the ha-ha and the paradox.

The 'Ha Ha', the only words Bosse-da-Nage ever utters in Faustroll, "is the idea of duality, of echo, of distance, of symmetry, of greatness and duration, of the two principles of good and evil." (Hugill 2012b) Referring to the yes-but statement Hugill says "this may be taken as a standard pataphysical response to any proposition (including this one).” And most obviously the antimony can be seen in all the contradictions that pataphysics is so fond of.

The antinomy, in a pataphysical sense, is the mutually incompatible or paradox. Mutually contradictory opposites can and do co-exist in the pataphysical universe.

3.2.3 Anomaly

The anomaly is the exception. And exceptions are important in pataphysics. But then again everything is equal, so in a pataphysical world no exceptions would exist at all, or rather, everything would be equally exceptional. The anomaly disrupts and surprises. Hugill mentioned a great example of a collection of anomalies: the sourcebook project by William Corliss, who collects scientific papers that are anomalous. Bök says it is “the repressed part of a rule which ensure that the rule does not work” (Bök 2002, p.38).

3.2.4 Sysygy

The syzygy surprises and confuses. It originally comes from astronomy and denotes the alignment of three celestial bodies in a straight line. In a pataphysical context it is the pun. It usually describes a conjunction of things, something unexpected and surprising. Serendipity is a simple chance encounter but the syzygy has a more scientific purpose. Bök mentions Jarry saying that the fall of a body towards a centre is the same as the ascension of a vacuum towards a periphery (Bök 2002, p.42).

A syzygy both surprises and confuses. The concept originally comes from the field of astronomy where it denotes the alignment of three celestial bodies. In a pataphysical context it usually describes a conjunction of things, something unexpected and surprising. Unlike serendipity, a simple chance encounter, the syzygy has a more scientific purpose. A typical instance is the pun, which Jarry called the syzygy of words (Jarry 1996). Next to being intentionally funny, puns demonstrate a clever use (or abuse) of grammar, syntax, pronunciation and/or semantics, often taken to a quite scientific level, such that without understanding of what is said and what is the intended meaning, the humour of the pun might be lost.

3.2.5 Clinamen

The clinamen is the unpredictable swerve that Bök calls “the smallest possible aberration that can make the greatest possible difference” (Bök 2002, p.43). He links it to Lucretius idea of an atom serving in its streamlined flow to create matter and to Epicurus’ parenklisis. But he also points out similarities to ideas like the Situationists’ ’détournement’, the reuse of pre-existing aesthetic elements and Hugill links it to the Dadaists’ ready-mades and Oulipo’s verbal games. An obvious example is Jarry’s ’merdre’, a swerve of the French word for shit (merde).

The concept of the clinamen can be understood as an unpredictable swerve which Bök called the smallest possible aberration that can make the greatest possible difference (Bök 2002). One of the most famous examples of a clinamen is Jarry’s merdre (the first word in his Ubu plays). He squeezed an extra ’r’ into the French word merde (meaning shit) and translates into something like pshit.

3.2.6 Absolute

The absolute is a reference to a transcended reality. Jarry talks about ’ethernity’ in Faustroll (Jarry 1996, p.104).

Others

Other concepts that are pataphysical or can be linked to it in a sense are alchemy and quantum mechanics. Alchemy because of its laws or equivalence and the union of opposites (Hugill 2012b) and quantum mechanics because of principles of uncertainty, indeterminacy and the idea of the multiverse of course.

Because string theory is speculation based on ideas that are themselves speculative (i.e., theories ofgeneral relativity and quantum mechanics), string theory is not in fact physics, but ’pataphysics.

Likewise, string theory and quantum calculations are, increasingly, not descriptive of an actual reality, but are simply mathematical pataphors. (P. Lopez) [4]

[4] http://www.urbandictionary.com/define.php?term=pataphysics

3.3 Unconscious

3.3.1 Oulipo

Finish section here. references and all

Potential literature is "the search for new forms and structures that may be used by writers in any way they see fit." Raymond Queneau (p2)

The Oulipo's goal is to discover new structures and to furnish for each structure a small number of examples. François Le Lionnais (p3)

"a formal quest" Warren Motte (p3)

Erecting the aesthetic of formal constraint, then, the Oulipo simultaneously devalues inspiration. (p10)

Three levels in the hierarchy of constraints:
1 Minimal level: constraints on the language in which the text is written
2 Intermedaite level: constraints on genre and certain literary norms
3 Maximum level: consciously preelaborated and voluntarily imposed systems of artifice
Oulipo is in the maximum level. François Le Lionnais (p11)

Oulipian systems of formal constraint are often based on the alphabet. François Le Lionnais (p13)

The nature of Oulipoan constraint is mathematical. François Le Lionnais

The Oulipo is anti-chance Claude Berge (p17)

> What is the objective of our work? To propose new 'structures' to writers, mathematical in nature, or to invent new artificial or mechanical procedures that will contribute to literary activity: props for inspiration as it were, or rather, in a way, aids for creativity. Raymond Queneau (p51)

> Aleatoricism is the incorporation of chance into the process of creation, especially the creation of art or media. The word derives from the Latin word alea, the rolling of dice. It should not be confused with either improvisation or indeterminacy.[1] Wikipedia

3.3.2 Borges

You could argue that by reading other literature that is pataphysical (whether or not it was intended) one can learn something about pataphysics. Reading Borges (Borges 2000, 1964, 1999; Borges and Guerrero 1957; Borges and Dembo 2010; Borges 2010) is a good example. His text "The analytical language of John Wilkins" (Borges 2000) contains a brilliant example of pataphysical thinking and coincidentally a good example of the kinds of search results my search tool should hopefully produce.

Referring to a certain Chinese dictionary entitled "The Celestial Emporium of Benevolent Knowledge" he claims that animals can be divided into:

1. those belonging to the Emperor
2. those that are embalmed
3. those that are tame
4. pigs
5. sirens
6. imaginary animals
7. wild dogs
8. those included in this classification
9. those that are crazy-acting
10. those that are uncountable
11. those painted with the finest brush made of camel hair
12. miscellaneous
13. those which have just broken a vase
14. those which, from a distance, look like flies

This kind of categorisation has also been discussed by Foucault in his book 'The Order of Things' (Foucault 1966).

Chapter 4

Creativity

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

---

finish intro to creativity and computers chapter

Educational psychologist Richard Mayer identified several different approaches to human creativity research and each approach has its own methodologies and goals. (Mayer 1999, p.453)

Psychometric
(creativity as a mental trait): quantitative measurement, controlled environments, ability based analysis

Psychological
(creativity as cognitive processing): controlled environments, quantitative measurements, cognitive task analysis

Biographical
(creativity as a life story): authentic environments, qualitative descriptions, quantitative measurements

Biological
(creativity as a physiological trait): physiological measures

Computational
(creativity as a mental computation): formal modelling

Contextual
(creativity as a context-based activity): social, cultural and evolutionary context

There are many debates across the involved disciplines. Specifically, Mayer identified five big questions of human creativity research: (Mayer 1999, p.450-451)

1. Is creativity a property of people, products, or processes?
2. Is creativity a personal or social phenomenon?
3. Is creativity common or rare?
4. Is creativity domain-general or domain-specific?
5. Is creativity quantitative or qualitative?

> An important challenge for the next 50 years of creativity research is to develop a clearer definition of creativity and to use a combination of research methodologies that will move the field from speculation to specification. (Mayer 1999, p.459)

> In research communities, approaches to the study of creativity differ in three main respects: 1) the type of research design, whether experimental, psychometric, observational etc. 2) the focus of the research, whether on human attributes cognitive processes or features of creative outcomes, and 3) the type of information that is used for the basis of evidence, by which is meant whether the time frame is present (real-time observation) or past (historical data) and whether the situation is artificial (laboratory) or natural (real world settings). (Candy 2012, p.3)

process
product

Creativity is a human quality - a process. Computers are essentially a product of human creativity.

Models of human creativity don't necessarily apply to computer creativity.

Creativity can be studied at various **levels** (neurological, cognitive, and holistic/systemic), from different **perspectives** (subjective and objective) and **characteristics** (combinational, exploratory and transformative).

This chapter introduces relevant models of human and computer creativity and describes the disciplines of computational creativity and creative computing.

These two simple statements already point to one of the main problems with evaluating creative computer software: do we evaluate the process or the product? See § 10.

put summaries at back of chapter or front? styling?

Summary:

- novelty/typicality/acceptability/variety/imagination/originality
- quality/value/appreciation/appropriateness/usefulness/relevance (/surprising?)
- efficiency/skill
- subjective/P/little-c
- objective/H/Big-C
- combinational, exploratory and transformative
- product/process
- The 4 P's
- Unified theory
- Associative and bisociative thinking
- Creative triptych (humour, discovery, art)
- 4 step model
- Problem solving
- P-creativity and H-creativity

4.1 In Humans

general introduction about human creativity models

Let us define creativity as *the ability to use original ideas to create something new and surprising of value*. We generally speak of creative ideas rather than products, since creative products merely provide evidence of a creative process that has already taken place.

> ‘Creativity is the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context’ (Plucker et al., 2004, p. 90) (Jordanous and Keller 2012)

4.1.1 Mel Rhodes and Ross Mooney

Mel Rhodes (1916-1976), who has a background in education and psychology, identified four common themes of creativity in 1961, which he termed “the four P’s of creativity” (Rhodes 1961):

Persons
: personality, intellect, temperament, physique, traits, habits, attitudes, self-concept, value systems, defence mechanisms and behaviour.

Process
: motivation, perception, learning, thinking and communication.

Press
: relationship between human beings and their environment

Products
: a thought which has been communicated to other people in the form of words, paint, clay, metal, stone, fabric, or other material.

Rhodes highlights the importance of a holistic view on creativity through these four areas of study, which he hoped would become the basis of a unified theory of creativity.

Where, what, who and how – those are the questions we need to ask regarding creativity.

Ross Mooney identified four aspects of creativity in 1963 (as cited in (Sternberg 1999)) which are essentially the same.

1. The creative environment
2. The creative person
3. The creative process
4. The creative product

Mel Rhodes vs Ross Mooney??

I believe these four aspects represent four conditions of creativity in a way, since all of them influence the outcome. Sometimes the creative environment and the creative person are merged into one, simply because people are often a product of their environment. The person will always be influenced by its surroundings, its culture, family, etc. and the environment alone cannot influence anything if not through a person. Another interpretation of course is that the creative environment is the context in which the creative product exists. Depending on which way we look at it, the first and second can be interchanged. This project focuses on the creative product, making sure the essence of the search tool is creative – the algorithms that define the main functionality of the search tool.

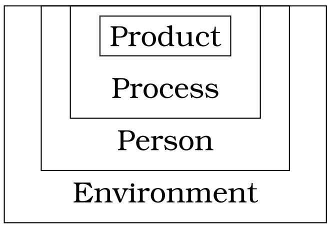

Figure 4.1: 4 Aspects of Creativity

Figure 4.1 shows how these aspects relate to each other. The environment influences all others and the person creates the product in a process.

4.1.2 Arthur Koestler

Arthur Koestler (1905-1983) published his study on creativity entitled "The Act of Creation" in 1964 (Koestler 1964). The book still carries influence today. His main contribution to the field is probably the concept of bisociation, a term he coined for the idea of two "self-consistent but habitually incompatible frames of reference" intersecting to give rise to new creative idea (Koestler 1964, p.35). It is interesting however to look at some of his other views on creativity as well.

He splits creativity into three domains, a triptych, without sharp boundaries: humour, discovery and art (see table 4.1). All creative acts traverse the three domains of this triptych from left to right, that is, the emotional climate of the creator changes “from an absurd through an abstract to a tragic or lyric view of existence” during the process (Koestler 1964, p.27). Central to all three domains is the “discovery of hidden similarities”, or bisociation. Koestler differentiates between associative thinking and bisociative thinking. He links those broadly to habit and originality, respectively. More specifically, associative thinking is conscious, logical, habitual, rigid, repetitive and conservative and bisociative thinking is unconscious, intuitive, original, flexible, novel and destructive/constructive.

Humour	$\Rightarrow$	**Discovery**	$\Rightarrow$	**Art**
Laugh		Understand		Marvel
Riddle		Problem		Allusion
Debunking		Discovering		Revealing
Coincidence		Trigger		Fate
Aggressive		Neutral		Sympathetic

Table 4.1: Koestler: Creative Triptych

4.1.3 Henri Poincaré, Graham Wallas and George Pólya

Henri Poincaré (1854-1912) (Poincare 2001) and Graham Wallas (1858-1932) (Wallas 1926) have defined a popular model (Koestler 1964; Boden 2003; Partridge and Rowe 1994) of the creative process (it was suggested by Poincaré ((Poincare 2001) book: “science and method”, chapter III: “mathematical discovery”, pages 387-400) and formulated by Wallas).

1. Preparation – focusing the mind on the problem
2. Incubation – unconscious internalising
3. Illumination – eureka moment from unconsciousness to consciousness
4. Verification – conscious evaluation of the idea and elaboration. . .

Weisberg criticises the stages of incubation and illumination (referred to by Partridge and Rowe 1994), saying that the creative process is really just simple problem solving, and that incubation is what he calls “creative worrying”.

> First, we have to **understand** the problem; we have to see clearly what is required. Second, we have to see how the various items are connected, how the unknown is linked to the data, in order to obtain the idea of the solution, to make a **plan**. Third, we **carry out** our plan. Fourth, we **look back** at the completed solution, we review and discuss it. (Polya 1957, p.5-6, his emphasis)

4.1.4 James Kaufman and Ron Beghetto

DOB of authors?

James C. Kaufman (1974-) and Ronald A. Beghetto (DOB?)...(See Kaufman and Beghetto 2009).

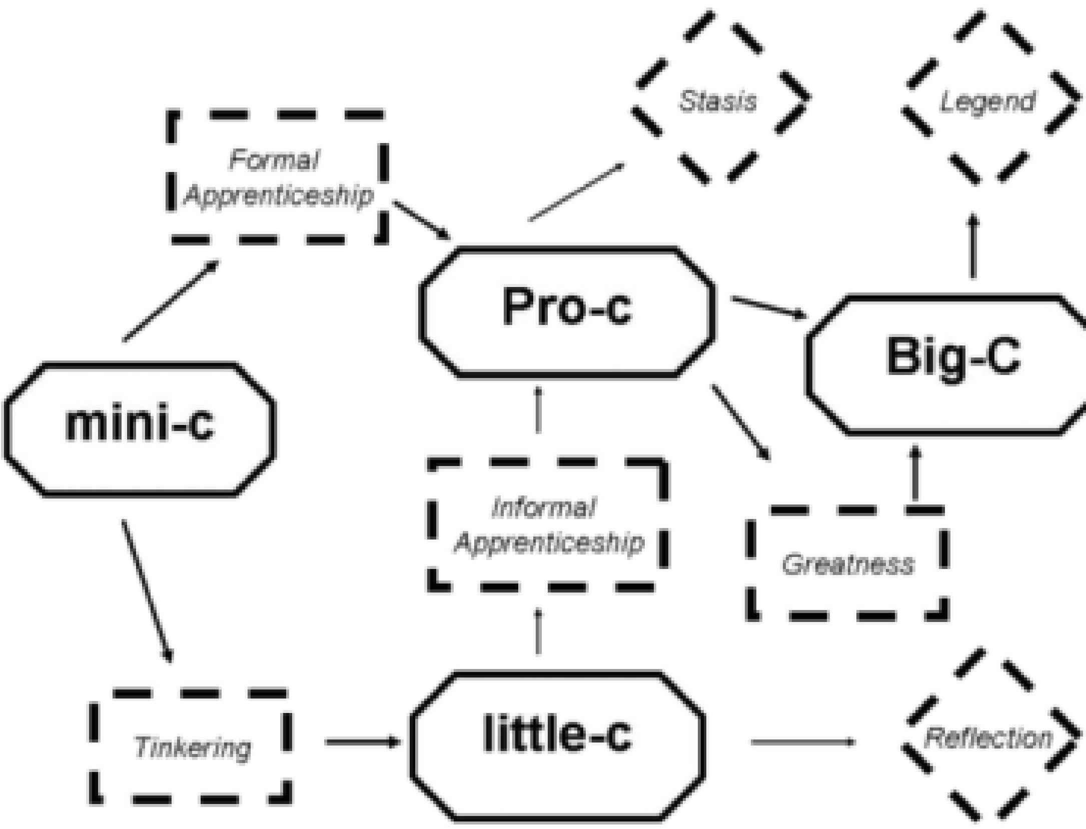

Figure 4.2: The 4 C Model

redo diagram

Big-C

Eminent Accomplishments. Big-C creativity consists of clear-cut, eminent

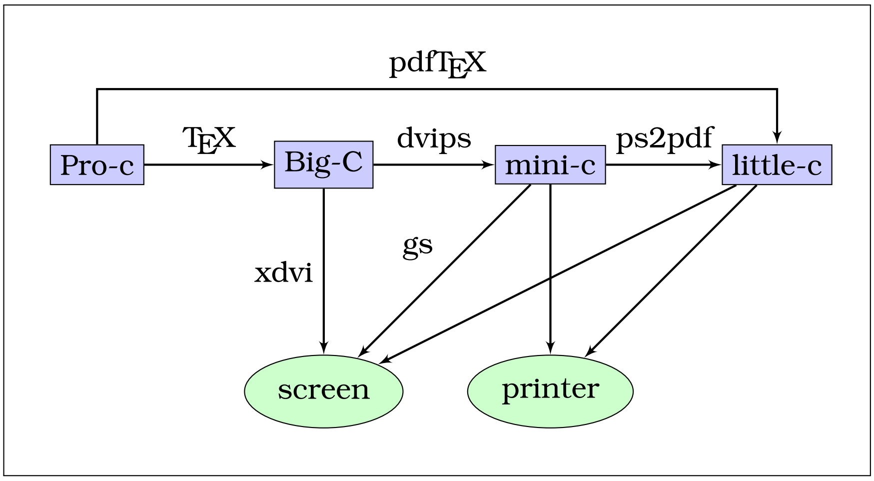

Figure 4.3: The 4 C Model2

creative contributions. Big-C creativity often requires a degree of time. Indeed, most theoretical conceptions of Big-C nearly require a posthumous evaluation.

Pro-c

Professional Expertise. Pro-c represents the developmental and effortful progression beyond little-c. The concept of Pro-c is consistent with the expertise acquisition approach of creativity.

ref

Ericsson1996, Ericsson2007 Propulsion Theory of Creative Contributions (Sternberg 1999, 2006): Replication, redefinition, forward incrementation, advance forward incrementation. Redirection, Reconstruction, reinitation, integration.

Little-c

Everyday Innovation. More focused on everyday activities, such as those creative actions in which the non-expert may participate each day.

Mini-c

Transformative Learning. Encompasses the creativity inherent in the learning process. “Mini-c is defined as the novel and personally meaningful interpretation of experiences, actions, and events.” (Beghetto and Kaufman 2007) Central to the definition of mini-c creativity is the dynamic, interpretive process of constructing personal knowledge and understanding within a particular sociocultural context. “a transformation or reorganization of incoming information and mental structures based on the individual’s characteristics and existing knowledge” **[p.63]Moran2003**

ref

Moreover, mini-c stresses that mental constructions that have not (yet) been expressed in a tangible way can still be considered highly creative.

> Mini-c highlights the intrapersonal, and more process focused aspects of creativity.

Applies to all: openness to new experiences, active observation, and willingness to be surprised and explore the unknown.

4.1.5 Margaret Boden

Professor Margaret Boden (1936-) is a prominent figure in the fields of Creative Computing (CC) and computational creativity. She has a background in medical sciences, psychology and philosophy and currently works as a cognitive scientist in computer science and artificial intelligence. Her main interest is in how the human mind works and how computer models of the mind and specific thinking processes can help us understand both better. She has provided two important contributions to the field. The first is her description of three distinct forms of creativity and the second is her important distinction between two senses of creativity (Boden 2003).

> [Creativity is] the ability to come up with ideas or artefacts that are **new, surprising and valuable**. (Boden 2003) (her emphasis).

She identified three distinct forms or cognitive processes of how creativity can happen. These are combinational, exploratory and transformational creativity, which can happen at the same time. (Boden 2003)[17, 21].

Combinational creativity
: making unfamiliar combinations of familiar ideas; juxtaposition of dissimilar; bisociation; deconceptualisation

Exploratory creativity
: exploration of conceptual spaces; noticing new things in old spaces

Transformative creativity
: transformation of space; making new thoughts possible by altering the rules of old conceptual space

Central to these three forms is the idea of a **conceptual space**. For any idea, its conceptual space describes the characteristics and constraints that define it in its most fundamental way. The conceptual space of a tea cup would contain information like: it is a container that can hold a hot fluid, it should hold about a half a pint of fluid and it might or might not be built in such a way as to not burn the hand that carries it. The specific colour of the cup or what material it is made of for example are not contained in its conceptual space.

Combinational creativity is the most common form of the three and is concerned with the unusual juxtaposition of common ideas. This aspect is highlighted in her definition of creativity, which requires novelty and surprise. The main idea is that any particular combination of ideas has to be unusual, causing surprise, but not (necessarily) the individual ideas themselves. She safeguards against purely random combination by including the usefulness of the result as a requirement in the definition. Exploratory creativity requires a person (or computer program) to fully explore the conceptual space of an idea and find unusual or interesting aspects of it. This form of creativity is about pushing an idea to its limits. Transformational creativity takes this exploration one step further. Once the limits of an idea have been identified, they can be transformed. This means that we can step out of the normal conceptual space of an idea, create a new one, alter or ignore the given constraints, add new ones, etcetera.

Boden argues that creative ideas are surprising because they go against expectations (Boden 2003). She also believes that constraints support creativity and are even essential for it to happen.

> Constraints map out a territory of structural possibilities which can then be explored, and perhaps transformed to give another one. (Boden 2003)

These three forms of creativity can be then interpreted on two levels. Any idea should be viewed and evaluated at the appropriate level. Consider the following scenario. A child and a professional architect both build a corbelled arch out of material available to them. Who is being creative here? The level of expertise is clearly different between the two. The child has no experience and is experimenting with the possibilities and limitations of the building blocks (exploring their conceptual space) while the architect has studied the technique for years and is simply applying knowledge he has learned from others (familiar use of a familiar idea). Clearly the child is being more creative in this example. Boden proposed to view and judge the creativity of these two persons separately by differentiating between two levels of creativity, a personal one and a historical one. **Psychological creativity** (P-creativity) is a personal kind of creativity that is novel in respect to an individual and **historical creativity** (H-creativity) is fundamentally novel in respect to the whole of human history. The child in the earlier scenario was P-creative but the architect was neither, he was simply applying his trained skills.

> P–creativity involves coming up with a surprising, valuable idea that's new to the person who comes up with it. It doesn't matter how many people have had that idea before. But if a new idea is H–creative, that means that (so far as we know) no one else has had it before: it has arisen for the first time in human history. (Boden 2003)

> Boden suggests that it is helpful to regard aspects such as novelty, quality and process as dimensions of creativity. Instead of asking 'is x creative?' (assuming a boolean judgement) or 'how creative is x?' (assuming a linear judgement) we should ask 'where does x lie in creativity space?' (assuming an n-dimensional space for n criteria where we can measure each dimension). (Pease et al. 2001, p.8)

> Boden argues that process does matter, stating that a program is creative only if it produces items in the right way - by transforming the boundaries of a conceptual space. This, she claims, can only be done if the program contains reflexive descriptions which mark its own procedures and is capable of varying them. The program should contain a meta-level which assesses methods of transforming a space and considers when and how to apply them. (Pease et al. 2001, p.8)

4.1.6 Robert J. Sternberg, James C. Kaufman and Timothy Leary

Sternberg and Kaufman identify a set of personality traits that are associated with creative people in their "Handbook of Creativity" (Sternberg 1999,?). These are independence of judgement, self-confidence, and attraction to complexity, aesthetic orientation, and tolerance for ambiguity, openness to experience, psychoticism, risk taking, androgyny, perfectionism, persistence, resilience, and self-efficacy. It is easy to find common characteristics among creative people but that doesn't mean that these automatically make a person or a product they make creative.

Timothy Leary took this idea of common characteristics a bit further and suggested there are four types of creative personalities ([25] as cited in [27]).

ref

From his ideas we can draw the conclusion that a creative person needs to be able to make novel combinations from novel ideas.

Reproductive Blocked
(no novel combinations, no direct experience)

Reproductive Creator
(no direct experience, but crafty skill in producing new combinations of old symbols)

Creative Creator
(new experience presented in novel performances)

Creative Blocked
(new direct experience expressed in conventional modes)

Tables 4.2 and 4.3 are in Leary's words.

Reproductive Blocked	**Reproductive Creator**	**Creative Creator**	**Creative Blocked**
The routine, well-socialised person who experiences only in terms of what he has been taught and who produces only what has been produced before.	The innovating performer who experiences only in terms of the available categories but has learned to manipulate these categories in novel combinations.	The person who experiences directly outside the limits of ego and labels, and who has learned to develop new models of communications, or who can manipulate familiar categories in novel combinations or who can let natural modes develop under his nurture.	The person who experiences uniquely and sensitively outside of game concepts (either by choice or helplessly by inability) but who is unable to communicate or uninterested in communicating these experiences outside the conventional manner.
Reproductive Performer	Creative Performer		Reproductive Performer
Reproductive Experience		Creative Experience	

Table 4.2: Leary's four types of creativity

<table>
<tr><th>Reproductive Blocked</th><th>Reproductive Creator</th><th>Creative Creator</th><th>Creative Blocked</th></tr>
<tr><td>Unimaginative, incompetent hack.</td><td>Reliable nihilist, insensitive, unsuccessful innovator whose shock value changes to morbid curiosity as fads of performance change.</td><td>The mad creative genius, the undiscovered far-out crackpot creator who is recognised by later generations as a creative giant.</td><td>Psychotic, religious crank, eccentric who uses conventional forms for expressing mystical convictions.</td></tr>
<tr><td>Competent, responsible, reliable worker.</td><td>Bold initiator who wins game recognitions but whose fame crumbles as fads of performance change.</td><td>The truly creative giant recognised by his own age and the ages to come.</td><td>Solid, reliable person with a “deep streak”.</td></tr>
<tr><td>Reproductive Performer</td><td colspan="2">Creative Performer</td><td>Reproductive Performer</td></tr>
<tr><td colspan="2">Reproductive Experience</td><td colspan="2">Creative Experience</td></tr>
</table>

Table 4.3: Leary’s social labels to describe the types of creativity

4.2 In Computers

In this section I am summarising a few models that try to implement creative thinking models in computers. It is really just a survey of different concepts and views and does not immediately apply to my specific research on creative search tools unfortunately.

4.2.1 Bipin Indurkhya

Indurkhya argues that there are two main cognitive mechanisms of creativity: namely juxtaposition of dissimilar and deconceptualization. He says that we are constraint by associations of our concept networks that we inherit and learn in our lifetime, but that computers do not have those conceptual associations and have therefore an advantage when it comes to creative thinking (Indurkhya 1997). He suggests a computer model using two layers that interact with each other: a perceptual and a conceptual layer.

- Juxtaposition of dissimilar
- Deconceptualization

4.2.2 Partridge and Rowe

Partridge and Rowe have written a good survey of computational models of creativity in their book “Computers and Creativity” (Partridge and Rowe 1994) although it is now probably quite out of date (the book was published in 1994). They mention the computer as an unbiased medium for executing creative programs (Partridge and Rowe 1994, p.26). Some of the computational methodologies they discuss are as follows, many taken from classical artificial intelligence research.

- Generative grammars
- Discovery programs
- Rule based systems
- Meta-rules (which reason about and create new rules)
- Analogical mechanisms
- Flexible representations
- Classifier systems
- Decentralised systems

- Connectionist systems
- Neural networks
- Emergent memory models

Classifier systems for example, consist of a set of rules and a message list.

1. Place input messages on current message list
2. Find all rules that can match messages
3. Each such rule generates a message for the new message list
4. Replace current message list with the new one
5. Process new list for any system output
6. Return to step 1

These can easily be combined with genetic algorithms to enable the system to learn an appropriate classifier set. This is called emergent behavior. Another approach is connectionism a.k.a. neural networks. They then go on to describe their emergent-memory model. They are applying the ideas of Poincare and Wallas and are heavily influence by Minsky's theory of K-lines (Minsky 1980, 1988). They define the following characteristics for creative programs:

- flexible knowledge representation scheme
- representational imprecision
- multiple representations
- self-assessment
- full elaboration

4.2.3 David Gelernter

Gelernter introduces a theory of how the human mind works in (Gelernter 1994). His "spectrum model" is based on the idea of mental focus and relates well to creativity. According to him we have a thought spectrum. The higher the mental focus, the more awake we are, the more adult we are and modern, logical and rational, convergent, abstract and detailed. The less focused we are the younger or ancient or dreaming we are. Low focus thoughts are metaphoric, hallucinations, divergent, creative, inspirations, concrete, ambient and emotional. Emotions glue low focus thoughts together.

He gives a good example of his own computer program that is being trained by a set of simple pairs (or memories) in the form -mood: happy- for example. These sets of pairs form the experience of the system, the memory that the system can access. It's fetching all memory pairs that match a certain probe, then generalizes them and picks out a feature that is common to all and then uses that to probe further if necessary.

He models his spectrum concept in a way that if we want the system to operate at low focus, more memory pairs would be fetched and more generalised features are deducted and so on. He describes his FGP program (Fetch Generalise Project) as follows (Gelernter 1994, p.132).

1. Fetch memory pairs in response to a probe (question)
2. Sandwhich them together and peer through the bundle at once
3. Notice the common features that emerge strongly (generalise)
4. Pick out interesting emergent details and probe further if necessary

With low focus the system would not generalise as much and just pick out a particular memory, etc. The computer system he has built seems very limited. His memory pairs cannot describe everything. For example they can describe states but not actions.

This idea of accessing thoughts/memories is very closely related to searching. Searching an index in a search engine is similar to remembering, trying to find all memories related to the current thought for example.

4.2.4 Marvin Minsky

Minsky introduces the concepts of k-lines in his Society of Mind (Minsky 1980, 1988). It is basically a theory of memory. He claims that the "function of a memory is to recreate a state of mind". His theory of k-lines is as follows.

> When you get an idea, or solve a problem, or have a memorable experience, you create what we shall call a K-line. This K-line gets connected to those mental agencies that were actively involved in the memorable mental event. When that K-line is later activated, it reactivates some of those mental agencies, creating a partial mental state resembling the original. (Minsky 1980, 1988)

This theory works quite well with Gelernter's idea of memory. K-lines in this sense are nothing other than Gelernter's memory pairs.

He and his student Push Singh have formalised the idea of a panalogy, which could be relevant for my project. The idea is that an idea can and should be conceptualised in many different ways. This could be seen as a fall-back mechanism for computational models, if one approach didn't return the desired/expected results.

4.2.5 Matthew Elton

Elton explains the concept of "Artificial Creativity" which can be seen as a subarea of Artificial Intelligence (AI). AI research isn't human enough, he argues, it needs to include less abstract ideas like emotions, morals, aesthetic sensibility and creativity. He goes on to explain in detail how production, evaluation and etiology play a role in everything (Elton 1995).

Opposed to the tradidtional approach of AI to study some aspect of the human brain in a specific domain only, he argues that in order to understand creativity we need to look at more than that. Creativtiy arises from a process that is not isolated. The etiology (its history) is essential for something to be classed as creative. Generation (of artefacts or ideas) cannot count as creative if it doesn't undergo evaluation in the process. In order to evaluate we need a sound knowledge of the relevant domain. "We want creative evaluation to be influenced by a longstanding history of interaction with entities (of whatever kind) in the world." Computer systems can be seen in two perspectives: plastic and implastic (resettable). "All systems can be seen from the implastic perspective since ultimately all systems are built out of physical components that are (statically) well behaved, but for certain explanatory purposes some are best understood plastically." Connectionist networks are an example of a plastic system. The brain is a plastic system too.

How do we get enough cultural information and background into the machine to train it? "There is no pure science of creativity, because it is paradigmatically idiographic - it can only be understood against the backdrop of a particular history."

His comments on evaluation are inspirational. How do I make my system evaluate its results or productions (as opposed to me testing my system)?

4.3 In Academia

Two transdisciplinary fields of study have emerged from the variety of disciplines concerned. These are computational creativity and creative computing. The former lies at the cross section of artificial intelligence and cognitive science and the latter is mostly distinguished by its involvement in art. Creative computing focuses on the process of creativity rather than just the outcome as in computational creativity.

Summary
- Boden: Combinational, exploratory and transformative (Boden 2003; Wiggins 2006) (process)
- Boden: new, surprising, valuable (Boden 2003) (product)
- Colton: Skill + appreciation + imagination = creativity (or the appearance of) (Colton 2008a) (product+process)
- Wiggins: relevance + acceptatbility + quality (Wiggins 2006) (product)
- Ritchie: typicality + quality (Ritchie 2001, 2007) (product)
- Pease: novelty + value (Pease et al. 2001) (product+process)
- Ventura: efficiency + variety (Ventura 2008) (product+process)
- Jordanious: value (related concepts: usefulness, appropriateness, relevance) + novelty (related concepts: originality, newness) (Jordanous and Keller 2012)

references

The concept of creative computing has existed for some time but has not yet managed to evolve into a recognised discipline within computer science. Computational creativity, on the other hand, has emerged as a field within artificial intelligence research [1] and overlaps with creative computing ideas to some extent.

It is important to differentiate between the terms creative computing and computational creativity. Intuitively the former is about doing computations in a creative way, while the latter is about achieving creativity through computation. You can think of the latter falling into the artificial intelligence category (using formal computational methods to mimic creativity as a human trait, see also [2]) and the former being a more poetic endeavour of how the computing itself is done, no matter what the actual purpose of the program is.

As a good example of creative computing, consider the International Obfuscated

[1] http://www.computationalcreativity.net/iccc2013/
[2] http://www.computationalcreativity.net/iccc2013/

C Code Contest [3]. The competition revolves around writing compilable/runnable code, while visually appearing as obfuscated as possible. They value unusuality, obscurity and creativity but expect contestants to follow the strict rules and constraints of the C programming language.

Examples of computational creativity are Simon Colton's Painting Fool [4] or Harold Cohen's AARON [5]; both are computer programs that paint pictures. Kurzweil's Cybernetic Poet [6] is a classic example of a program that produces poetry.

But how may we apply the insights into creativity described above in computing? One approach is described by Simon Colton (Colton 2008b), who suggests we should adopt human skill, appreciation and imagination.

> Without skill, they would never produce anything. Without appreciation, they would produce things which looked awful. Without imagination, everything they produced would look the same. (Colton 2008b)

He thinks that evaluating the worth of an idea or product is the biggest challenge facing computational creativity. Whereas in conventional problem solving success is defined as finding a solution, in a creative context more aesthetic considerations have to be taken into account. He suggests three ways for computer programs to generate creative artefacts:

1. Mimicking human skill
2. Mimicking human appreciation
3. Mimicking human imagination

4.3.1 Computational Creativity

Computational creativity is a relatively new discipline and as such not well defined. Simon Colton, the creator of the Painting Fool, describes it as the discipline of generating artefacts of real value to someone (Colton 2008b). This is in contrast to classic artificial intelligence problem solving. He identifies that evaluating the worth of such an artefact as the biggest problem of computational creativity. In problem solving, success is when a solution to the problem

[3] http://www.ioccc.org/
[4] http://www.thepaintingfool.com/
[5] http://www.kurzweilcyberart.com/aaron/history.html
[6] http://www.kurzweilcyberart.com/poetry/rkcp_overview.php

has been found. In artefact generation a more aesthetic consideration has to be taken into account.

One could say that computational creativity is the attempt at giving computers the skills, appreciation and imagination needed to produce creative artefacts. Whether or not this makes the computer creative or the programmer is another question that I will not try to answer here.

Computational creativity has emerged from within AI research. Simon Colton and Geraint Wiggins argue AI falls within a problem solving paradigm: "an intelligent task, that we desire to automate, is formulated as a particular type of problem to be solved " (Colton and Wiggins 2012, p.2), whereas "in Computational Creativity research, we prefer to work within an artefact generation paradigm, where the automation of an intelligent task is seen as an opportunity to produce something of cultural value. " (Colton and Wiggins 2012, p.2, my emphasis)

The International Association for Computational Creativity (ACC) promotes the advancement of computational creativity which is defined as follows.

> Computational Creativity is the art, science, philosophy and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative. (International Conference on Computational Creativity (ICCC)14 website)

Computational creativity is multidisciplinary, bringing together researchers from artificial intelligence, cognitive psychology, philosophy, and the arts. Its role within computer science falls under the scientific paradigm (Hugill 2013, p.8), (see also Eden 2007), as opposed to CC in the technocratic paradigm. Its main goal is to model, simulate or replicate human creativity using a computer and it has the following three aims:

- to construct a program or computer capable of human-level creativity
- to better understand human creativity and to formulate an algorithmic perspective on creative behavior in humans
- to design programs that can enhance human creativity without necessarily being creative themselves

The ACC manages the annual ICCC, whose recent call for papers (for ICCC 2014) gives a useful insight into their research agenda. It can be broken down as follows:

- Paradigms, metrics, frameworks, formalisms, methodologies, perspectives
- Computational creativity-support tools
- Creativity-oriented computing in education
- Domain-specific vs. generalised creativity
- Process vs. product
- Domain advancement vs. creativity advancement
- Black box vs. accountable systems

Simon Colton and Geraint Wiggins have also identified several directions for future research in the field: (Colton and Wiggins 2012, p.5)

1. Continued integration of systems to increase their creative potential.
2. Usage of web resources as source material and conceptual inspiration for creative acts by computer.
3. Using crowd sourcing and collaborative creative technologies bringing together evaluation methodologies based on product, process, intentionality and the framing of creative acts by software.

This reminds of the 4 P's, and CC and Digital Humanities (DH) models

- Domain-specific vs. generalised creativity
- Process vs. product
- Domain advancement vs. creativity advancement
- Black box vs. accountable systems

4.3.2 Creative Computing

rewrite and format

In the recent first issue of the International Journal of Creative Computing (IJCrC) Hugill and Yang introduced CC as a new discipline (Hugill and Yang 2013) with an overarching theme of "unite and conquer" (Yang 2013, p.1, his emphasis). Its broad aim is to "reconcile the objective precision of computer systems (mathesis) with the subjective ambiguity of human creativity (aesthesis)." (Hugill and Yang 2013, p.5). Hugill and Yang suggest CC falls within the technocratic paradigm of computing (see also Eden 2007, p.8), i.e. the discipline is closest related to software engineering, rather than mathematics or natural sciences. They identify five main topics for CC research (Hugill and Yang 2013, p.15-17):

Challenges
: transdisciplinarity, cross-compatibility, continuity and adaptivity

Types
: creative development of a product, development of a CC product and development of tool for creativity support

Mechanisms
: Boden's combinational, exploratory and transformational creativity

Methods
: development of suitable transdisciplinary CC research methodologies

Standards
: resist standardisation, novel, continuous user interaction, creative mechanisms

The main challenge is for technology to become "more adaptive, smarter and better engineered to cope with frequent changes of direction, inconsistencies, irrelevancies, messiness and all the other vagaries that characterise the creative process" (Hugill and Yang 2013, p.5). In part, these issues are due to the transdisciplinary nature of the field and factors such as common semantics, standards, requirements and expectations are typical challenges. Hugill and Yang therefore argue that creative software should be flexible and able to adapt to ever changing requirements, it should be evaluated and re-written continuously and it should be cross-compatible.

The different **types** of CC highlight the different aspects researcher and practitioners focus on during their work. These are

Process
: creative development of a computing product,

Product
: development of a Creative Computing product and

Community
: development of computing environment to support creativity.

The creative computing process should consist of combinational, exploratory and transformational activities (in the sense of Margaret Boden's theory, as discussed in cross ref

).

Broadly speaking, you could say that approaches to CC are therefore either bottom-up (1) or top-down (2).

The third type of CC in a way reflects what Hugill and Yang call the "local and global levels", which represent the two types of creativity identified by Boden (P- and H-creativity, see above). It is concerned with developing environments, tools and methods and the management of these.

This includes cross-compatibilty, which directly represents the solution to the personal/local and historical/global issues mentioned by Boden and Hugill and Yang.

Similar to the four step model of the creative process by Poincaré and Wallas (Poincare 2001; Wallas 1926) and the four step model of problem solving by Pólya (Polya 1957), they propose a four step model for the creative computing process. They do this by comparing the acts of artistic creation and software engineering in some detail. They found that the two processes follow essentially the same levels of abstraction (from the abstract to the concrete). The four steps are (Hugill and Yang 2013, p.15):

1. Motivation (digitised thinking)
2. Ideation (design sketch)
3. Implementation (creative system)
4. Operation (effect of system/revision)

This reminds of the 4 P's, and CC and DH models ??

Given the transdisciplinary nature of CC, Hugill and Yang suggest that existing research methodologies are unsuitable and new ones have to be developed. The following is an example of a possible CC research methodology they propose as a starting point (Hugill and Yang 2013, p.17):

1. Review literature across disciplines
2. Identify key creative activities
3. Analyse the processes of creation
4. Propose approaches to support these activities and processes
5. Design and implement software following this approach
6. Experiment with the resulting system and propose framework

Hugill and Yang propose four **standards** for CC (Hugill and Yang 2013, p.17) namely, resist standardisation, perpetual novelty, continuous user interaction and combinational, exploratory and or transformational.

Summary
- Transdisciplinary
- Technocratic paradigm of computer science
- Mathesis + aesthesis
- Local + global
- Top-down + bottom-up
- Continuous life-cycle, cross-compatibility , adaptive software, interoperability

4.3.3 Speculative Computing

SpecLab (Drucker 2009) is a book by Johanna Drucker about her experiences as a researcher moving between disciplines and the projects she worked on as part of the Digital Humanities laboratory at the University of Virginia, USA. Several of those had pataphyscial inspirations.

In his review, on the back cover of the book, John Unsworth says that Drucker "emphasizes the graphical over the textual, the generative over the descriptive, and aesthetic subjectivity over analytical objectivism." Her main argument is that in the design of digital knowledge representation, subjectivity and aesthetics are an essential feature. She confronts logical computation with aesthetic principles with the idea that design is information.

Aesthesis is the theory of ambiguous and subjective knowledge, ideological and epistemological, while Mathesis is formal objective logic and they contrast each other. Knowledge is always interpretation and subjectivity is always in opposition to objectivity. Knowledge becomes synonymous with information and as such can be represented digitally as data and metadata.

> Arguably, few other textual forms will have greater impact on the way we read, receive, search, access, use and engage with the primary materials of humanities studies than the metadata structures that organize and present that knowledge in digital form. (Drucker 2009, p.9)

But how is this metadata analysed? How do we analyse this type of structured data? And most important of all she asks, what can be considered as data, what can be expressed in those quantitative terms or other standard parameters? Is data neutral, raw or does it have meaning? Here she also points out that many information structures have graphical analogies and can be understood as diagrams that organize the relations of elements within the whole.

Because "computational methods rooted in formal logic tend to be granted more authority [. . .] than methods grounded in subjective judgement", she introduces the discipline of Speculative Computing as the solution to that problem. The concept can be understood as a criticism of mechanistic, logical approaches that distinguish between subject and object.

> Speculative computing takes seriously the destabilization of all categories of entity, identity, object, subject, interactivity, process, or instrument. In short, it rejects mechanistic, instrumental, and formally logical approaches, replacing them with concepts of autopoiesis (contingent interdependency), quantum poetics and emergent systems, heteroglossia, indeterminacy and potentiality, intersubjectivity, and deformance. Digital Humanities is focused on texts, images, meanings, and means. Speculative Computing engages with interpretation and aesthetic provocation. (Drucker 2009, p.29)

Pataphysics governs exceptions and anomalies and she introduces a, what she calls, "patacritical" method of including those exceptions as rules - even if repeatability and reliability are compromised. Bugs and Glitches are privileged over functionality, and although that may not be as useful in all circumstances, they are "valuable to speculation in a substantive, not trivial, sense." In an essay on speculative computing (Drucker and Nowviskie 2007) she says "Pataphysics celebrates the idiosyncratic and particular within the world of phenomena, thus providing a framework for an aesthetics of specificity within generative practice." To break out of the formal logic and defined parameters of computer science we need speculative capabilities and Pataphysics. "The goal of pataphysical and speculative computing is to keep digital humanities from falling into mere technical application of standard practices."

> 'Pataphysics inverts the scientific method, proceeding from and sustaining exceptions and unique cases, while quantum methods insist on conditions of indeterminacy as that which is intervened in any interpretative act. Dynamic and productive with respect to the subject-object dialectic of perception and cognition, the quantum extensions of speculative aesthetics have implications for applied and theoretical dimensions of computational humanities. (Drucker and Nowviskie 2007)

With this, Drucker introduces Speculative Aesthetics, which links interface design in which other speculative computing principles. She also refers to Kant and his idea of "purposiveness without purpose." She says that the appreaciation of design as it is (outside of untility) is the goal of speculative aesthetics.

4.3.4 Digital Humanities

Anne Burdick, Johanna Drucker, Peter Lunefeld, Todd Presner and Jeffrey Schnapp (referred to as 'the authors' in this section) have collaboratively written an authoritative manifesto for the field of DH ((Burdick et al. 2012). Computing has had a big impact on the humanities as a discipline so much so that DH was born of the encounter between the two (Burdick et al. 2012, p.3). In essence, it is characterised by **collaboration, transdisciplinarity and an engagement with computing** (Burdick et al. 2012, p.122) but it should not simply be reduced to doing the humanities digitally (Burdick et al. 2012, p.101). It spans across many traditional areas of research, such as literature, philosophy, history, art, music, design and of course computer science.

Transliteracy[7] therefore is fundamental (Thomas et al. 2007);

"The field of Digital Humanities may see the emergence of polymaths who can 'do it all"': who can research, write, shoot, edit, code, model, design, network, and dialogue with users. (Burdick et al. 2012, p.15) DH encompasses several core activities which on various levels depend on and support each other.

Design
: Shape, scheme, inform, experience, position, narrate, interpret, remap/reframe, reveal, deconstruct, reconstruct, situate, critique

Curation, analysis, editing, modelling
: Digitise, classify, describe, metadata, organise, navigate

Computation, processing
: Disambiguate, encode, structure, procedure, index, automate, sort, search, calculate, match

Networks, infrastructure
: Cultural, institutional, technical, compatible, interoperable, flexible, mutable, extensible

Versioning, prototyping, failures
: Iterate, experiment, take-risks, redefine, beta-test

[7] Sue Thomas et al. define transliteracy as "the ability to read, write and interact across a range of platforms, tools and media from signing and orality through handwriting, print, TV, radio and film, to digital social networks." (Thomas et al. 2007)

IF THE STUDY OF ART OR HUMAN CREATIVITY FALLS WITHIN HUMANITIES RESEARCH, THEN COMP CREAT SHOULD FALL WITHIN DIGITAL HUMANITIES, RIGHT, AND USE THE TOOLS AND METHODS AVALIBALE.

DESIGN

The authors suggest that “for digital humanists, design is a creative practice harnessing cultural, social, economic, and technological constraints in order to bring systems and objects into the world.” (Burdick et al. 2012, p.13)

In generative mode, these designers shape structural logics, rhetorical schemata, information hierarchies, experiential qualities, cultural positioning, and narrative strategies. When working analytically, their task is to visually interpret, remap or reframe, reveal patterns, deconstruct, reconstruct, situate, and critique. (Burdick et al. 2012, p.12)

CURATION, ANALYSIS, EDITING, MODELING

> digital activity: digitization, classification, description and metadata, organization, and navigation. (Burdick et al. 2012, p.17)

> Involving archives, collections, repositories, and other aggregations of materials, CURATION is the selection and organization of materials in an interpretive framework, argument, or exhibit. (Burdick et al. 2012, p.17)

> The parsing of the cultural record in terms of questions of authenticity, origin, transmission, or production is one of the foundation stones of humanistic scholar- ship upon which all other interpretive work depends. But editing is also productive and generative, and it is the suite of rhetorical devices that make a work. Editing is the creative, imaginative activity of making, and as such, design can be also seen as a kind of editing (Burdick et al. 2012, p.18)

> MODELING highlights the notion of content models—shapes of argument expressed in information structures and their design. (Burdick et al. 2012, p.18)

COMPUTATION, PROCESSING

> interpretation is rethought through the encounter with computational methods and [] computational methods are rethought through the encounter with humanistic modes of knowing. (Burdick et al. 2012, p.103)

> 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)

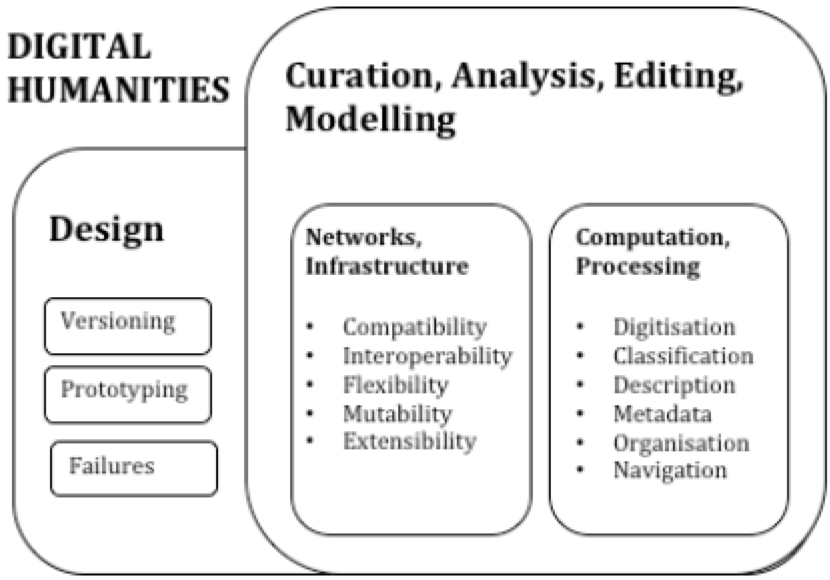

Figure 4.4: Digital Humanities model

NETWORKS, INFRASTRUCTURE

> Designing and building digital projects depend on knowledge of these fundamentals and on a nuanced understanding of the networked environments in which the projects will develop and variously reside. (Burdick et al. 2012, p.17)

> Digital work takes place in the real world, and humanists once accus- tomed to isolated or individualized modes of production must

> now grapple with complex partnerships and with insuring the long-term availability and viability of their scholarship (Burdick et al. 2012, p.21)

VERSIONING, PROTOTYPING, FAILURES

> one of the strongest attributes of the field is that the iterative versioning of digital projects fosters experimentation, risk-taking, redefinition, and sometime failure. (Burdick et al. 2012, p.21)

SOUNDS LIKE SOFTWARE ENGINEERING

> It is important that we do not short-circuit this experimental process in the rush to normalize practices, standardize methodologies, and define evaluative metrics. (Burdick et al. 2012, p.21)

argument for creative computing too

Field map of digital humanities: emerging methods and genres

(Burdick et al. 2012, p.29-60)

```
• enhanced critical curation
o digital collections
o multimedia critical editions
o object-based argumentation
o expanded publication
o experiential and spatial
o mixed physical and digital
• augmented editions and fluid textuality
o structured mark-up
o natural language processing
o relational rhetoric
o textual analysis
o variants and versions
o mutability
• scale: the law of large numbers
o quantitative analysis
```

o text-mining

o machine reading

o digital cultural record

o algorithmic analysis

• distant/close, macro/micro, surface/depth

o large-scale patterns

o fine-grained analysis

o close reading

o distant reading

o differential geographies

• cultural analytics, aggregation, and data-mining

o parametrics

o cultural mash-ups

o computational processing

o composite analysis

o algorithm design

• visualization and data design

o data visualization

o mapping

o information design

o simulation environments

o spatial argument

o modelling knowledge

o visual interpretation

• locative investigation and thick mapping

o spatial humanities

o digital cultural mapping

o interconnected sites

o experimental navigation

o geographic information systems (GIS)

o stacked data

• the animated archive

o user communities

o permeable walls

o active engagement

o bottom-up curation

o multiplied access

o participatory content creation

• distributed knowledge production and performative access

o global networks

o ambient data

    - collaborative authorship
    - interdisciplinary teams
    - use as performance
    - crowd-sourcing
- humanities gaming
    - user engagement
    - rule-based play
    - rich interaction
    - virtual learning environments
    - immersion and simulation
    - narrative complexity
- code, software, and platform studies
    - narrative structures
    - code as text
    - computational processes
    - software in a cultural context
    - encoding practices
- database documentaries
    - variable experience
    - user-activated
    - multimedia prose
    - modular and combinatoric
    - multilinear
- repurposable content and remix culture
    - participatory Web
    - read/write/rewrite
    - platform migration
    - sampling and collage
    - meta-medium
    - inter-textuality
- pervasive infrastructure
    - extensible frameworks
    - heterogeneous data streams
    - polymorphous browsing
    - cloud computing
- ubiquitous scholarship
    - augmented reality
    - web of things
    - pervasive surveillance and tracking
    - ubiquitous computing
    - deterritorialization of humanistic practice

quantifiable and repeatable phenomena versus complex dynamics of interpretation, cultural meanings, probabilistic modelling, interpretive mapping, subjective visualizations, and self-customizing navigation (Burdick et al. 2012, p.103)

TOOLS

> Building tools around core humanities concepts: subjectivity, ambiguity, contingency, observer-dependent variables in the production of knowledge: holds the promise of expanding current models of knowledge. As such, the next generation of digital experimenters could contribute to humanities theory by forging tools that quite literally embody humanities centred views regarding the world. (Burdick et al. 2012, p.104)

> 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)

> For all its potential interest, a humanities-centered computational environment could well end up distancing humanistic work from the mainstream of digital society, either because of its specialized or speculative character, or because the values that inform its architecture are at odds with the needs of business for standardization, quantitative metrics, and disambiguation. (Burdick et al. 2012, p.105)

Summary
- Collaborative, Transdisciplinary and Computing

4.3.5 Computer Ethics

ETHICS: PROCESS< PRODUCT< PURPOSE
ROBOT ETHICS: similar to 4-p's of creativity (McBride 2013)

it has three actors: Robot engineer, client and user.

4 approaches:
- challenge the myth of autonomy

• Developing practice-based approaches (in context of it purpose and environment)
• Managing ethical variety
• A model for human0centred robot ethics

Virtuous robot:
• Human-centred
• Man-machine interdependency
• Practice based (context)
• Ethical variety

Chapter 5

Technology

Parts of this chapter have been published in peer-reviewed...

Knowledge needed to understand project:

- Search engines
- index
- corpus
- query - expansion etc
- results
- searching vs browsing
- web programming
-

update all graphics with inkscape

5.1 Searching vs. Browsing

put section into foundations chapter?

What do we actually mean by 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 users' point of view the search process can be broken down into four activities (Sutcliffe and Ennis 1998) reminiscent of classic problem solving techniques (Polya 1957):

Problem identification
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 when we consider a situation 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" (Marchionini and Shneiderman 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**) (de Vries 1993; Marchionini 2006). A situation where an IN cannot easily be articulated or in fact is not existent (the user is not looking for anything specific) can be considered a typical case of exploratory search and describes the kind of search that is most suited to our proposed tool. 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 certain extend of interpretation or prediction (e.g. analysis, evaluation, forecasting, transformation).

When it comes to exploratory search though, 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 exploring (Marchionini 2006).

> Much of the search time in learning search tasks is devoted to examining and comparing results and reformulat-ing 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 (de Vries 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).

All of these ideas, however, seem to be concerned with how users interact with a search system, rather than how the system acts itself. So we need to shift our perspective and think about how a search tool can be more supportive for exploratory search directly and by what means.

5.2 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)

In simple terms, a typical search process can be described as follows. 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.

Most big 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 (Google 2012) that influence the ranking of web pages including their original PageRank algorithm (Brin and Page 1998a,b). We can only speculate whether these signals also take into account some creative factors due to their secrecy. Other search engines like YossarianLives (currently in alpha release) [1] concentrate on purely abstract concepts like metaphors for their search algorithms.

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. 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).

IR does not only refer to the retrieval of online information from the Internet but generally also for any offline databases for example. In this survey I will focus on IR for the World Wide Web and whenever I'm speaking of an IR system I will generally refer to a web search engine system unless stated otherwise. The architecture of a search engine is made up of different components so to speak. A crawler, an indexer and a ranking system and the user interface. See figure 5.1.

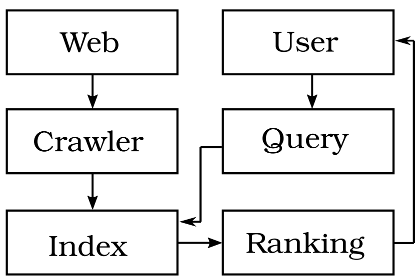

Figure 5.1: Abstract search engine architecture

[1] http://www.yossarianlives.com/

5.2.1 IR Models

IR models describe ranking algorithms formally. ???

There are different models for different needs, for example a multimedia system is going to be different than a text based 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 or 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.

The notation for IR models is as follows (quoted from Baeza-Yates and Ribeiro-Neto 2011, p.58):

An IR model is a quadruple $[D, Q, F, R(q_i, d_j)]$ where:

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

5.2.2 The Boolean Model

The similarity of document d_j to query q is defined as follows (quoted from (Baeza-Yates and Ribeiro-Neto 2011, p.65))

$$sim(d_j, q) = \begin{cases} 1 & \text{if } \exists\ c(q) \mid c(q) = c(d_j) \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

Sometimes things are not quite black and white though and we need to weigh the importance of words somehow. The easiest way to do that is by looking at the frequency in which a word occurs.

5.2.3 The Vector Model

The vector model allows a more flexible scoring since it basically computes the various degrees of similarity between documents (taken from (Baeza-Yates and Ribeiro-Neto 2011, p.78)).

$$\begin{aligned} \vec{d_j} &= (w_{1,j}, w_{2,j}, \dots, w_{t,j}) \\ \vec{q} &= (w_{1,q}, w_{2,q}, \dots, w_{t,q}) \end{aligned} \tag{5.2}$$

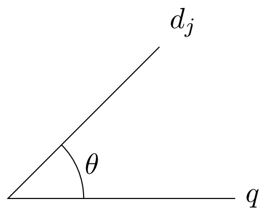

Figure 5.2: The Vector Model

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.

$$\begin{aligned} sim(d_j, q) &= \frac{\vec{d_j} \cdot \vec{q}}{|\vec{d_j}| \times |\vec{q}|} \\ &= \frac{\sum_{i=1}^{t} w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^{t} w_{i,j}^2} \times \sqrt{\sum_{i=1}^{t} w_{i,q}^2}} \end{aligned} \tag{5.3}$$

Here is an example algorithm for computing this score taken from (Manning et al. 2009, p.125).

Where,

q	is the query
N	is the total number of documents
d	is a document
t	is a query term
wt_q	is the weight of the term in the query
tft_d	is the term frequency of t in d

```
CosineScore(q)
  float Scores[N] = 0
  for each d
  do Initialise Length[d] to the length of document d
  for each query term t
  do calculate wt,q and fetch postings list for t
    for each pair(d, tft,d) in postings list
    do add wft,d to Scores[d]
  Read the array Length[d]
  for each d
    do Divide Scores[d] by Length[d]
    return Top K components of Scores[]
```

Figure 5.3: Pseudo-code for computing vector scores

wft_d	is the $tf - idf$ weight of t in d
K	is the number of results we want
$postingslist$	is the list of all (d, tft_d) for a given t.

There are several other common IR models that I won't discuss in detail here. These include the probabilistic, set-based, extended Boolean and fuzzy set (Widyantoro and Yen 2001; Miyamoto 2010, 1988; Srinivasan 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).

Total Term Frequency

$$F_i = \sum_{j=1}^{N} f_{i,j} \tag{5.4}$$

Where F_i is the total frequency of term k_i in the collection and $f_{i,j}$ is the frequency of occurance of term k_i in document d_j and N is the total number of documents.

Term Frequency TF

normalised using log function

$$tf_{i,j} = \begin{cases} 1 + \log f_{i,j} & \text{if } f_{i,j} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

Inverse Document Frequency IDF

$$idf_j = \log \frac{N}{df_i} \tag{5.6}$$

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.

TF-IDF Weighting

$$w_{i,j} = \begin{cases} (1 + \log f_{i,j}) \times \log \frac{N}{df_i} & \text{if } f_{i,j} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

Where $w_{i,j}$ is the weight associated with (k_i, d_j). In simpler notation this is:

$$w_{i,j} = tf_{i,j} \times idf_j \tag{5.8}$$

This means that rare terms have a higher weight and more so if they occur a lot in one document.

Architecture

Index

describe index

An index or inverted index is blah blah blah.... diagram?

A 'conjunctive component' describes which terms occur in a document and which ones do not. E.g. for $V = \{k_1, \dots, k_t\}$, if $[k_1, k_2, k_3]$ occur in d_j then it could be $(1, 1, 1)$ which means all terms occur, or $(1, 0, 0)$ which means only term k_1 appears in d_j.

$c(d)$	is the term conjunctive component for document d
$c(q)$	is the term conjunctive component for query q

The **term-document matrix** can be written as follows, where $f_{i,j}$ is the frequency of term k_i in document d_j.

$$\begin{array}{c} \\ k_1 \\ k_2 \\ k_3 \end{array} \begin{array}{c} \begin{matrix} d_1 & d_2 \end{matrix} \\ \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{bmatrix} \end{array} \tag{5.9}$$

This is how an index could look like (all possible words that occur within a collection of documents), but usually an inverted index is prepared for further calculations. This inverted index is basically a list of words (the so called dictionary or vocabulary) with each word pointing to a list (the so called postings list) that indicates the documents in which the word occurs. The inverted index can include other values such as the document frequency (the number of documents that contain the term).

The dictionary is usually processed to eliminate punctuation and stop words (e.g. a, and, be, by, for, the, on, etc.) that would be useless in everyday text search engines. Techniques that are useful for this are the following (Manning et al. 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

For specific domains it even makes sense to build a 'controlled vocabulary' which can be seen as a domain specific taxonomy and are very useful for query expansion.

Search Algorithms

5.2.4 Ranking

Ranking signals 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.

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 [2] for instance) would be much higher and 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. Here is a survey of the most notable algorithms.

PageRank was developed in 1998 by Larry Page and Sergey Brin as part of their Google search engine and announced in their often cited paper (Brin and Page 1998b) and they further describe the algorithm here (Brin and Page 1998a). 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).

$$PR(a) = \frac{q}{T} + (1-q)\sum_{i=1}^{n} \frac{PR(p_i)}{L(p_i)} \tag{5.10}$$

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

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

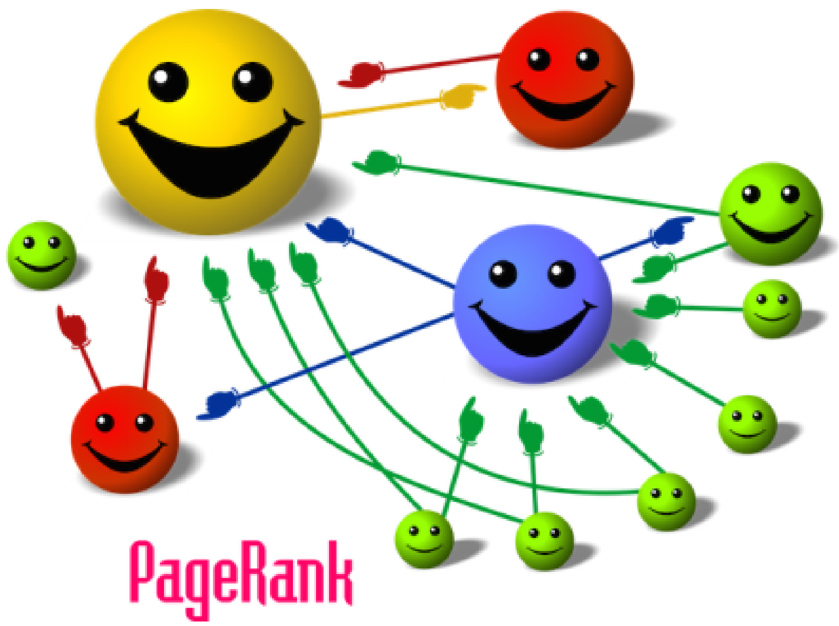

Figure 5.4: PageRank algorithm illustration from Wikipedia

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

The HITS algorithm also works on the links between pages. It was first described by Kleinberg (Kleinberg 1999; Kleinberg et al. 1999, p.472) in 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. Again, the following formula is taken from (Baeza-Yates and Ribeiro-Neto 2011, p.471). S is the set of pages.

$$\begin{aligned} H(p) &= \sum_{u \in S | p \to u} A(u) \\ A(p) &= \sum_{v \in S | v \to p} H(v) \end{aligned} \tag{5.11}$$

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 Mihaila in 2000 in (Bharat and 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.

Another algorithm is the so called Fish search algorithm. It was first described by De Bra in 1994 (De Bra and Post 1994a,b; De Bra et al. 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 et al. 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 et al. 2006), social annotations (Bao et al. 2007), trust (Garcia-molina et al. 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 et al. 2010; Taye 2009).

5.2.5 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 analysis 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.

Challenges of Web Search

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 et al. 2003).

Summary

IR refers to the retrieval of information from a collection. In terms of the Internet it is often called Web search. A Web search engine is divided into different components, being the crawler to build an index of the collection and a ranking algorithm which stands between the index and the user.

Different retrieval models exist including the Boolean and the Vector model. Other methods exist to make search results more accurate, including relevance feedback and query expansion.

Search quality is generally measured using the metrics of precision and recall but for Web search precision is more important and usually a metric called "precision at n" is used for measurements.

Challenges are the size of the World Wide Web and ambiguous, unstructured nature of Web pages among others.

Ranking can be done at different stages of the search process. Depending on how the index is formatted and what information can be pre-computed at that stage, the ranking algorithm evaluates every page for relevance and returns them in order. There exist lots of different approaches on ranking, including PageRank and HITS (both analyse the link structure of the WWW), or more dynamic models like Fish search or genetic approaches.

5.3 Natural Language Processing

describe NLTK and the core functionality

Natural Language Toolkit (NLTK) Python library[3].

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.

5.3.1 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

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

$d_{a,b}(|a|,|b|)$ where:

$$d_{a,b}(i,j)=\begin{cases}\max(i,j) & \text{if } \min(i,j)=0\\ \min\begin{cases}d_{a,b}(i-1,j)+1\\ d_{a,b}(i,j-1)+1\\ d_{a,b}(i-1,j-1)+1_{a_i\neq b_j}\\ d_{a,b}(i-2,j-2)+1\end{cases} & \text{if } i,j>1 \text{ and } a_i=b_{j-1} \text{ and } a_{i-1}=b_j\\ \min\begin{cases}d_{a,b}(i-1,j)+1\\ d_{a,b}(i,j-1)+1\\ d_{a,b}(i-1,j-1)+1_{a_i\neq b_j}\end{cases} & \text{otherwise.}\end{cases} \tag{5.12}$$

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.

Natural Language Processing (NLP) blah blah blah...

Bird, S., Klein, E. and Loper, E., 2009. NLP with Python 1st ed., Sebasopol, CA: O'Reilly Media.(Bird et al. 2009)

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

Taken from [1]:(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. . .

5.3.2 Damerau-Levensthein

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 $\rightarrow$ 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?

$$P(w \mid h) = \frac{count(hw)}{count(h)} \tag{5.13}$$

Using the **chain rule of probability**:

$$\begin{aligned} P(w_1^n) &= P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1^2)\ldots P(w_n \mid w_1^{n-1}) \\ &= \prod_{k=1}^{n} P(w_k \mid w_1^{k-1}) \end{aligned} \tag{5.14}$$

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

$$P(w_1^n) = \prod_{k=1}^{n} P(w_k \mid w_{k-1}) \tag{5.15}$$

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)

$$P(w_n \mid w_{n-N+1}^{n-1}) = \frac{C(w_{n-N+1}^{n-1} w_n)}{C(w_{n-N+1}^{n-1})} \tag{5.16}$$

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

$$PP(W) = \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P(w_i \mid w_{i-1})}} \tag{5.17}$$

Smoothing

Add-One : Laplace smoothing for bigrams

$$P_{Add-1}(w_i \mid w_{i-1}) = \frac{c(w_{i-1}, w_i) + 1}{c(w_{i-1}) + V} \tag{5.18}$$

Adjusted count

$$c_i^* = (c_i + 1)\frac{N}{N + V} \tag{5.19}$$

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.

$$c^* = (c + 1)\frac{N_{c+1}}{N_c} \tag{5.20}$$

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.

$$P(c \mid d, \lambda) = \frac{\exp \sum_i \lambda_i f_i(c, d)}{\sum_{c\prime} \exp \sum_i \lambda_i f_i(c\prime, d))} \tag{5.21}$$

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

$$\hat{c} = \underset{c \in C}{\text{argmax}}\ P(c \mid d, \lambda) \tag{5.22}$$

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

Table 5.1: MaxEnt Example table

Features:

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

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

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

$$empirical\ E(f_i) = \sum_{(c,d)\ \in\ observed(C,D)} f_i(c,d) \tag{5.23}$$

Evaluation

$$Precision = \frac{\text{number of correctly labeled}}{\text{total number of extracted}} \tag{5.24}$$

$$Recall = \frac{\text{number of correctly labeled}}{\text{total number of gold}} \tag{5.25}$$

$$F_1 = \frac{2PR}{P+R} \tag{5.26}$$

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 et al. 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/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]$^{\text{LOCATION}}$./.

$$t_1^n = \underset{t_1^n}{\operatorname{argmax}}\, P(w_1^n \mid t_1^n) P(t_1^n) \tag{5.27}$$

$$P(t_i \mid t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})} \tag{5.28}$$

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

$$P(\text{NN} \mid \text{DT}) = \frac{C(\text{DT}, \text{NN})}{C(\text{DT})} = \frac{56,509}{116,454} = 0.49 \tag{5.29}$$

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.

$$P(\text{is} \mid \text{VBZ}) = \frac{C(\text{VBZ}, \text{is})}{C(\text{VBZ})} = \frac{10,073}{21,627} = 0.47 \tag{5.30}$$

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:

Figure 5.5: Grammers

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)))
    (, ,)
    (NP
      (NP (DT the) (JJ Eighth) (NN day))
```

```
    (PP (IN of)
      (NP (NNP February) (, ,) (NNP Pursuant)))
    (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-))
    (, ,)
    (NP
      (NP (NNS proprietors))
      (PP (IN of)
        (NP
          (NP (DT a) (NN house) (NN situate))
          (PP (IN at)
            (NP (NNP Paris))))))
    (, ,)
    (NP (CD 100) (NN bis))
    (, ,))
  (VP (VBP rue)
    (NP
      (NP (NNP Richer))
      (, ,)
      (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))))
    (. .)))
```

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.

5.4 Linguistics / WordNet

Here's my hypernym term. holonym hypernym

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.

5.5 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

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

Chapter 6

Evaluation

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

Summary

- output – input
- novelty + value
- product + process
- Mimicking
- novelty + value + quality
- randomness + serendipity

reformat

Useless Search Results

The word useless is defined as “not fulfilling or not expected to achieve the intended purpose or desired outcome” in the Oxford dictionary (2010). Given this definition most of the search results we have in mind would be classed as useless. That is at least if we considered every result individually, outside of context and with an information-lookup query in mind. If we have an exploratory search in mind however things get more interesting and we will explain why in the remainder of the paper.

Relevant versus Creative

When we say relevant results we mean the kind of search results that any mainstream search engine would produce, the kind of results you would immediately understand their connection to the query for, the kind of results that just makes sense. Consider the example results in table 1.

Results which might seem useless at first can be much more creative or even poetic. And creative results support exploratory search. Surprise and user expectations play a big role in creativity according to Boden (2003).

Fewer expectations (an open mind) allow creativity to happen more easily. Empirical experiences form expectations, which hinder our ability to accept creative ideas when they happen. In order to be able to recognise creative ideas we need to be able to see what they all have in common and in what way they differ and not reject unusual, unexpected ones.

Relevant Creative

UK action sports lifestyle brand (www.animal.co.uk) List of animals in the Emperor's possession Wikipedia article (en.wikipedia.org/wiki/animal) Instructions about embalming animals Animal Planet - Discovery Channel (animal.discovery.com) Information on a society for animal training Table 1 - Results for a query for animal. The relevant results were retrieved from Google. The creative results were inspired by Borges' "Chinese Encyclopaedia" from The Analytical Language of John Wilkins (1942) We can link this very nicely to the idea of exploratory search. Lowering expectations or opening the mind implies extending the task domain or problem space. Creativity and exploratory search seem predestined to work with each other.

Biases

Wikipedia defines bias as "an inclination to present or hold a partial perspective at the expense of (possibly equally valid) alternatives. Anything biased generally is one-sided, and therefore lacks a neutral point of view." (2012)

However, biases can be good and bad. It is important to consider the implications of their existence though, especially when trying to measure the success of something objectively. An example of when biases can be advantageous is location signals that the search tool takes into account when producing results. An Englishmen would probably not have much use of a Chinese website and vice-versa, even if the actual content matches the original query (unless of course the user happens to understand both languages perfectly). Another example of this is location queries such as 'Chinese restaurants in Cambridge', which should return web pages about restaurants based in Cambridge, UK or Massachusetts, USA, depending on the user's I.P. address. This might seem logical, but in the truest sense it is a bias employed by the search engine to help provide more relevant results to individuals. Truly unbiased search results are probably impossible to come by nowadays.

There is a general move from objectivity to subjectivity in the sense that users become the subject of search results as much as the query they pose. Instead of neutrally providing results for a query alone, the results are tailored around the information known about the user (e.g. language, location, clickstream, social media likes, bookmarks, etc.) to make up the missing context. The user becomes the subject and context of a query, while the results become an objective list of matches for all those values rather than just the query term(s).

Standard Web search: Subject/User Object/Results

Constraints

There are certain factors and constraints that influence the perception and success of the results. Some can be taken into account when building a search system but others cannot be avoided. User education is one way to deal with those issues. Earlier we briefly mentioned some external constraints such as the setting in which the search takes place. Is the user operating from a handheld device or a desktop computer? Is he or she in a hurry to find answers or just leisurely browsing for them? Is the search system web-based or is the user querying a database?

User Expectations It is important to note that "search systems are not used in isolation from their surrounding context, i.e., they are used by real people who are influenced by environmental and situational constraints such as their current task" (White and Marchionini 2004). User expectations should be taken into consideration during the evaluation of search results. Users who are hoping to find precise answers to a specific question might not be satisfied by exploratory search results. Someone browsing for inspiration on a broad topic on the other hand could benefit from them. Users should therefore be informed about the nature of the search tool in some way.

User Skill The searching skills of the user matter. Specifically his or her ability to articulate an information need and any knowledge of special search techniques (use of Boolean modifiers, quotation marks, wildcards, etc.) are two important factors that influence the results obtained greatly. This is very much based on the old idea of garbage-in, garbage-out (Lidwell et al. 2010). Visual Representation The way that results are presented affects how the user perceives them. A diversity of different document types, for example text, images, sound, or video results could improve how well the results are rated (Sawle et al. 2011). Johanna Drucker had already pointed out that “many information structures have graphical analogies and can be understood as diagrams that organise the relations of elements within the whole” (2009). An alphabetical list is a typical model for representing text data sets for example. But is a ranked list really the best way to represent search results? Other models could be a differently ranked or ordered list, a tree structure, a matrix, a one-to-many relationship, etc.

Structure of Results As suggested by Sawle et al (2011) we need to consider different ways to structure and measure search results. A single, perfectly good result might be deemed irrelevant and useless if it is surrounded by several unsuitable results. Therefore there might be certain advantages to measuring and evaluating the value or relevance of individual results over a whole set of results.

Direct User Relevance Feedback Relevance feedback lets users rate individual results or sets of results either directly (through manual ratings) or indirectly (through click-stream data). This data is then congregated and used for webpage rankings or other purposes such as suggesting other query terms. It can improve results for similar queries in the future but also lets the user stir the direction his search is taking in real-time. Users can adjust their query to manipulate the results; this basically means they adjust some of their own constraints.

“Relevance feedback—asking information seekers to make relevance judgments about returned objects and then executing a revised query based on those judgments—is a powerful way to improve retrieval.” (Marchionini 2006)

Automatic Query Expansion As opposed to integrating and involving the user actively in the refinement of a query, in automatic query expansion the improvements are done passively, often completely without the user’s knowledge. Information gathering methods include, for example, the analysis of mouse clicks, so called like buttons (e.g. Facebook, Google+) or eye tracking, etc. How the collected data is then used varies. Simple examples of automatic query expansion are the correction of spelling errors or the hidden inclusion of synonyms when evaluating a query.

Depending on these factors and constraints, search results can be viewed as useful or useless. In a way the usefulness or correctness of an idea or result cannot always be judged fairly – there are always conditions that will affect how the outcome is interpreted. In the scenario of a creative search tool, results could be very useful, while they might be completely useless in another.

6.1 IR Evaluation

In this paper (Sawle et al. 2011) we have discussed an initial approach to measure the creativity of search results. Based on a definition of creativity by Boden, we attempt to define creativity in a way which could be applied to search results and provide a simple metric to measure it.

Evaluating search results is not easy. The most widely used metric to measure the quality of results is that of precision and recall.

Precision is defined as the fraction of retrieved documents that are relevant.

$$Precision = \frac{relevantdocumentsretrieved}{retrieveddocuments} \tag{6.1}$$

Recall is defined as the fraction of relevant documents that are retrieved.

$$Recall = \frac{relevantdocumentsretrieved}{relevantdocuments} \tag{6.2}$$

Note the slight difference between the two. Precision tells us how many of all retrieved results were actually relevant (of course this should preferable be very high) and recall simply indicates how many of all possible relevant documents we managed to retrieve. This can be easily visualised as follows.

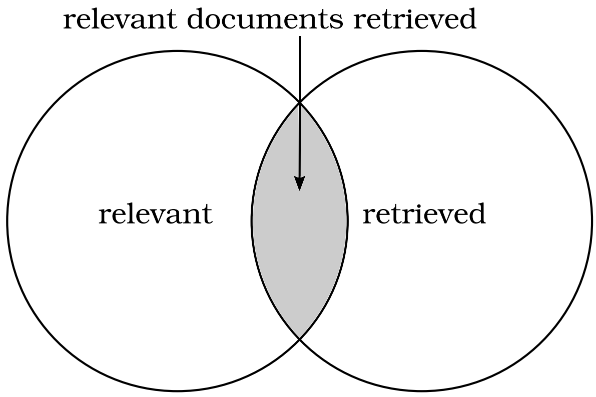

Figure 6.1: Precision and Recall

Precision is typically more important than recall in web search while it is the other way around in a database search system maybe. The mean average precision value (MAP) can be calculated following this formula (taken from (Baeza-Yates and Ribeiro-Neto 2011, p.141):

$$MAP_i = \frac{1}{|R_i|} \sum_{k=1}^{|R_i|} P(R_i[k]) \tag{6.3}$$

Where R_i is the set of relevant documents for query q_i.

But for many web searches is it not necessary to calculate the average of all results, since users don't inspect results after the first page very often and it is therefore desirable to have the highest level of precision in the first 5 to 30 results maybe. For this purpose it is common to measure the average precision of web search engines after only a few documents have been seen. This is called "Precision at n" or "P@n" (Baeza-Yates and Ribeiro-Neto 2011, p.140). So for example this could be P@5 or P@10 or P@20. For example, to compare two ranking algorithms, we would calculate P@10 for each of them over an average of 100 queries maybe and compare the results and therefore the performance of the algorithm.

The Text REtrieval Conference (TREC) is a conference that provides large test sets of data to participants and lets them compare results. They have specific test sets for web search comprised of crawls of .gov web pages for example, but unfortunately they have to paid for to get a copy. [1]

There are certain other factors that can be or need be evaluated when looking at the complete search system. For example the time it takes to crawl for data, the time it takes to index data, the amount of storage needed for data and then also how fast the query response is and how many queries the system can handle in a given time period.

6.2 Approaches

see §4

Taking the debates about human creativity and directly applying them to machines seems logical but may be the wrong and lazy approach. Adapting Mayer's five big questions to machines does not seem to capture the real issues at play.

[1] http://ir.dcs.gla.ac.uk/test_collections/

1. Is creativity a property of programmers, users, machines, products, or processes?
2. Is creativity a local, a network or an Internet phenomenon?
3. Is creativity common or rare? (P or H creativity)
4. Is creativity domain-general or domain-specific?
5. Is creativity quantitative or qualitative?

- Can a machine judge whether a human is creative?
- Is creativity a property of machines (hardware or software?)
- Is mimicking human creativity really enough and appropriate?
- Compare against "human creativity"? Or define machine creativity from scratch?
- Who is creative? The programmer or the program?
- Can creativity be objectively measured?
- Quantitative or qualitative?
- In respect to P or H creativity?
- Output minus input? (we don't have the same strict judgement on humans)
- Is it the product or the process or both?
- Does context matter? (Blind deaf dumb person = computer?)
- Does time matter?
- Does purpose or intention matter?
- AGI vs AI? Artificial general creativity vs artificial creativity?

On a more practical level, there are various problems that arise when trying to evaluate computer creativity. Anna Jordanous found that "evaluation of computational creativity is not being performed in a systematic or standard way" (Jordanous 2011, p.2, her emphasis).

Pease and Colton [27] divide it into two notions:

ref

- whether an idea or artefact is valuable or not, and
- whether a system is acting creatively or not.

6.3 Output Minus Input

It has been argued that "creativity may be seen as **output minus input**." (Pease et al. 2001, p.2, my emphasis). The output in this case is the creative product but the input is not the process. Rather, it is the **"inspiring set"** (comprised of explicit knowledge such as a database of information and implicit knowledge input by a programmer) of a piece of software.

> The degree of creativity in a program is partly determined by the number of novel items of value it produces. Therefore we are interested in the set of valuable items produced by the program which exclude those in the inspiring set. (Colton et al. 2001, p.3)

> We are interested in how a single agent can come up with something that is novel **relative to its initial state of knowledge** (Ritchie 2007, p.72, his emphasis)

"Output minus input" might easily be misinterpreted as "product minus process", however, that is not the case. In fact, Pease, Winterstein and Colton argue that "the process by which an item has been generated and evaluated is intuitively relevant to attributions of creativity" (Pease et al. 2001, p.6), and that "two kinds of evaluation are relevant; the evaluation of the item, and evaluation of the processes used to generate it." (Pease et al. 2001, p.7). If a machine simply copies an idea from its inspiring set then it just cannot be considered creative, and they need to be disqualified so to speak.

6.4 Measurable Attributes

Simon Colton came up with an evaluation framework called the "creative tripod". The tripod consists of three behaviours a system or artefact should exhibit in order to be called creative. The three legs represent **skill, appreciation, and imagination** and three different entities can sit on it, namely the programmer, the computer and the consumer. Colton argues that the perception "that the software has been skillful, appreciative and imaginative, then, regardless of the behaviour of the consumer or programmer, the software should be considered creative." (Colton 2008a, p.5) + (Colton 2008b, p.5). As such a product can be considered creative, if it appears to be creative. If not all three behaviours are exhibited, however, it should not be considered creative. (Colton 2008a, p.5) + (Colton 2008b, p.5)

> Imagine an artist missing one of skill, appreciation or imagination. Without skill, they would never produce anything. Without appre-

> ciation, they would produce things which looked awful. Without imagination, everything they produced would look the same (Colton 2008a)

Pease, Winterstein and Colton suggest that all creative products must be **novel and valuable** (Pease et al. 2001, p.1) and provide several measures that take into consideration the context, complexity, archetype, surprise, perceived novelty, emotional response and aim of a product. In terms of the creative process itself they only discuss **randomness** as a measurable approach. Elsewhere, Pease et al discuss using **serendipity** as an approach (Pease et al. 2013).

Davide Piffer suggests that there are three dimensions of human creativity that can be measured, namely **novelty, usefulness/appropriateness and impact/influence**. (Piffer 2012, p.258-259). As an example of how this applies to measuring a person's creativity he proposes citation counts (Piffer 2012, p.261). While this idea works well for measuring scientific creativity maybe, he does not explain how this would apply to a visual artist for example [2].

Graeme Ritchie identifies three main criteria for creativity as **novelty, quality and typicality** (Ritchie 2007, p.72-73), although he argues that "novelty and typicality may well be related, since high novelty may raise questions about, or suggest a low value for, typicality" (Ritchie 2007, p.73) (see also Ritchie 2001). He proposes several evaluation criteria which fall under the following categories: (Ritchie 2007, p.91-92) basic success, unrestrained quality, conventional skill, unconventional skill, avoiding replication and various combinations of those.

This also somewhat suggests that variety is a criteria for creativity.

Dan Ventura later suggested the addition of **variety and efficiency** to Ritchie's model (Ventura 2008, p.7).

Geraint Wiggins introduced a formal notation and set of rules for the description, analysis and comparison of creative systems (Wiggins 2006) which is largely based on Boden's theory of creativity (Boden 2003). The framework uses three criteria for measuring creativity: **relevance, acceptability and quality**.

Anna Jordanous proposed 14 key components of creativity, an ontology of creativity (Jordanous and Keller 2012, p.104-120), from a linguistic analysis of creativity literature which identified words that appeared significantly more often in discussions of creativity compared to unrelated topics. (Jordanous and Keller 2012, p.120).

[2] http://www.artfacts.net seems to provide just that though.

> The themes identified in this linguistic analysis have collectively provided a clearer 'working' understanding of creativity, in the form of components that collectively contribute to our understanding of what creativity is. Together these components act as building blocks for creativity, each contributing to the overall presence of creativity; individually they make creativity more tractable and easier to understand by breaking down this seemingly impenetrable concept into constituent parts. (Jordanous and Keller 2012, p.120)

The 14 components Jordanous collated are: (Jordanous and Keller 2012, p.118-120)

1. Active Involvement and Persistence
2. Generation of Results
3. Dealing with Uncertainty
4. Domain Competence
5. General Intellect
6. Independence and Freedom
7. Intention and Emotional Involvement
8. Originality
9. Progression and Development
10. Social Interaction and Communication
11. Spontaneity / Subconscious Processing
12. Thinking and Evaluation
13. Value
14. Variety, Divergence and Experimentation

6.5 Linda Candy

Evaluation is well established in HCI. HCI is not unsimilar to creativity. Design too.

I guess her work is meant mostly for "interactive art" while mine is meant for creative computing, but clearly there are many overlaps.

In HCI, historically, the focus has been on people as users deploying task oriented systems. The criteria for evaluation has largely been in terms of ease of use, task efficiency and effectiveness- usability. However, attributes such as speed and productivity are, for the most part, meaningless in the context of creative interactive experience. (Candy 2012, p.23)

Evaluation "is used to describe assessing and judging the value or worth of a particular idea or artifact both during the creative process and afterwards. Whether the process is systematic or ad hoc, evaluation depends upon criteria and measures that are situated and domain specific." (Candy 2012, p.7)

"Whatever the context, evaluation is always tailored to the approach, needs, purpose and methodology of that context." (Candy 2012, p.7)

"For evaluation to contribute to a successful outcome, the practitioner needs to have the necessary information including constraints on the options under consideration." (Candy 2012, p.7)

"The matrix for evaluating creativity represents three standpoints: the capabilities of the creator, the audience, or more accurately, participant, experiences, and the features of the interactive systems as artworks. This initial matrix has been extended to include creative processes for both creator and audience participant (i.e. working practices and interaction experiences) and contextual factors in the form of the physical and technical environment in which the creative acts and events take place, including the influence of physical and technical resources and real world constraints." (Candy 2012, p.7-8)

"Evaluation studies are well established in the field of Human-Computer Interaction (HCI) as well as interaction design contexts in general." (Candy 2012, p.8)

"The evaluation of user, or rather, participant, experience of interactive artworks often involves measurement of aesthetic appreciation and the various engagement qualities which are dependent on personal traits, motivations, expectations, emotions and cognitive states of the audience. Those experiences that involve open-ended activity tend towards the creative end of human activity and, as such, are hardly ever measurable in quantitative ways." (Candy 2012, p.8)

"Evaluation is a key activity in creative design that can be revealed through documentation from design rationale. The introduction of rationale has been an important contribution to the quest for clarity and traceability in design decision-making. Design rationale may be thought of as structured records of design that support the understanding of decisions taken and allow designers to give better informed reconsideration to them at a later stage." (Candy 2012, p.9)

"A software system can be viewed as an artifact that embodies implicit theoretical constructs that are realized as functional and operational requirements (Carroll and Campbell 1989). Structures are chosen because of their ability to achieve the intended functionality, and such choices may be evaluated against various criteria. During the design process, the ideas are modified and there is a clarification and refinement of intended functions and features. There may be additional factors arising from the context of the project that affect the way the design is carried out: for example the need to keep sight of general applicability whilst meeting the domain specific requirements, or the influence of the given hardware platform and software tools. Whatever the situation, the relationship between designers' decision making and the design outcome is not necessarily transparent and this is can be a problem when it comes to system maintenance. The explicit listing of decisions made during a design process, and the reasons why those decisions were made provides a means to record and communicate the reasoning and justification behind a design decision, including alternatives considered and constraints that affected the decision-making including why alternatives were rejected. The successful application of design rationale to software system design can provide a form of communication of intent from the designer to those who are to maintain the system." (Candy 2012, p.9)

"A promising approach to the externalization of decision-making during the design process is being explored within practice-based research in the creative arts in the form of documented reflective practice. The approach builds upon a normal part of creative practice whereby practitioners draw and note ideas, designs and options in their sketch and notebooks. In this way, the documentation of tentative ideas and how they are worked into firmer proposals through testing and evaluation is a familiar and integral part of creativity. In practice-based research, documented reflective practice and empirical studies are frequently brought together." (Candy 2012, p.10)

The Multi-dimensional Model of Creativity and Evaluation (MMCE) shown in Figure 1 has four elements: people, process, product and context. (Candy 2012, p.11)

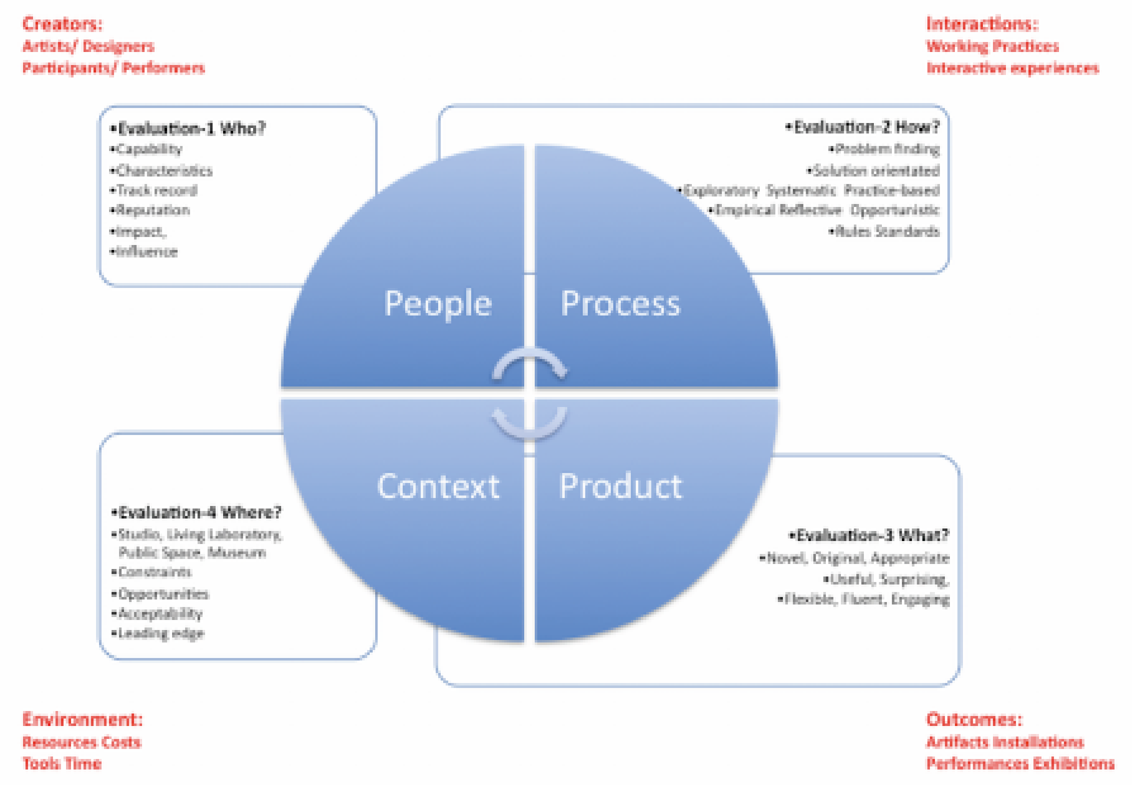

Figure 6.2: Candy's Multi-dimensional Model of Creativity and Evaluation

PEOPLE AS CREATORS: (Candy 2012, p.14-15)

Criteria for evaluating creator capability:

1. The creator must be able to demonstrate an ability to create an artistic outcome where subject matter, ideas and technique are combined well to produce a coherent outcome.
2. The creator must be able to demonstrate an ability to make work that is exploratory, creative and imaginative. Interesting ideas are presented in intelligent and surprising ways.
3. In respect of Composition and Interpretation, the creator must be able to demonstrate the ability to:
   - Select subject matter that is appropriate to a given theme
   - Manipulate ideas and techniques in a coherent manner
   - Express ideas visually (visual communication)
   - Respond in an individual and personal way

PROCESS AS INTERACTION: (Candy 2012, p.17)

Criteria for evaluation can be expressed as follows 1. For a work to be deemed

Categories in Creativity	Actors & Elements	Instances	Feature Descriptors	Outcomes	Criteria, Qualities, Values to Measure
People - Who is involved?	Creators	i) designer, artist	imagination, artistry, expertise, skill experience, intention, reputation	ii) participant, performer	profile, demographic, motivation, skills, experience, curiosity commitment
Process - How is it done?	Interactions	i) working practices	inspirational, solution driven, goal oriented, problem solver, systematic, exploratory, reflective, risk taking	ii) interactive behaviors	opportunistic, adventurous, curious, cautious, expert, knowledgeable, experienced
Product - what comes out?	Outcomes	i) artifacts, installations	novel, original, surprising, compelling, adaptable, aesthetically pleasing, effective	ii) performances, experiences	immediate, engaging, purposeful, enhancing, exciting, disturbing,
Context - where does it happen and with what?	Environments	i) studio, living laboratory ii) public spaces, museums	physical space, light, technical facilities, rules, costs, time resources, effort, organizational constraints/support	i) prototypes ii) exhibitions	design quality, usable, convincing, adaptable, effective, innovative, transcendent

Table 1: Categories of Creativity: Actors and Elements with Evaluation Measures: Criteria, Qualities, Value

Figure 6.3: Caption text under figure

engaging, the participant should exhibit observable responses. There are likely to be different levels of engagement depending on whether or not the audience has had prior experience of this kind of artwork or installation or similar.
2. The participant responses demonstrate active engagement in three ways: Immediate, Sustained or Creative. The categories are defined as follows:
• Immediate engagement: the work grabs immediate attention and yet is not so mundane as to create boredom.
• Sustained engagement: the work must excite curiosity in the and also be accessible to a general audience.
• Creative engagement: the work must excite immediate attention and encourage an audience to interact with it in a playful/purposeful way. As attention declines with familiarity and time, changes take place in the work that renew audience engagement.

PRODUCT AS OUTCOME: (Candy 2012, p.18)

Typical features for judging artworks include composition, aesthetic, affect, content, and technique. Criteria for evaluation can be expressed as follows:

• the composition of work should be coherent, exhibit shape and balance between order and complexity.
• the work should exhibit outstanding visual and sound qualities in color, line and form.
• the work should be pleasing, challenging, exciting, original etc.
• the content should be appropriate and effective for the chosen subject matter
• the execution should demonstrate high quality technique that fits the form.

It is interesting, therefore, to consider how criteria for judging the digital arts are specified by the Prix Ars Electronica, an international competition for Cyber Arts and the foremost event of its kind today.

Entries are judged by a Jury of experts in the order of their arrival and according to the following categories:

• Aesthetics • Originality
• Excellence of execution
• Compelling conception
• Innovation in technique of the presentation

CONTEXT AS ENVIRONMENT: (Candy 2012, p.21)

Establishing a workable living laboratory for interactive art and evaluation involved setting down acceptance criteria for assessing whether or not a new interactive art system was ready to be deployed.

These included: • degree of robu
stness of the art system in expectation of heavy public use • appropriate accessibility in respect of type of audience (e.g. children)
• adherence to safety and house rules required by the museum
• impact of other coinciding exhibits (sound, noise, light impacts)
• attention to participant orientation and training
• attention to art system maintenance by creator and technical support

> "we need to apply strategies for generating clear and unambiguous data that can be turned into meaningful information. From meaningful information, we can then derive understandings related to the context of use, the outcome of which might take the form

of a coherent model." (Candy 2012, p.21)

"Observation as a method for data collection raises issues as to its reliability in creativity evaluation. Data from observing creativity depends upon the interpretation of what the individual observer sees." (Candy 2012, p.22)

"However, in order to ‘measure’ creativity, we have to conduct research outside of controlled laboratory conditions, and cannot rely on fixed criteria that can be applied to all cases. The shifting ground and the ever-changing contexts often renders consistency out of reach." (Candy 2012, p.22)

"If the term ’measurement’ does not match what we are doing within the creativity domain, then why do we still use this word?" (Candy 2012, p.22)

"Whether an action is successful or unsuccessful depends on whether the intended result is achieved." (Candy 2012, p.23)

"Measuring success is more likely to be dependent on factors such as whether or not the system has engaged the audience in a playful or immersive way or whether it has elicited curiosity or excitement or concentrated attention and so on." (Candy 2012, p.23)

Part III

THE CΘRE: TΣCHNΘ-LΘGIC

Chapter 7

Theoretical Foundations

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

Why pataphysics? How does pataphysics relate to creativity and how does it support creativity in computers?

- link the various creativity models
- how do they apply to computers?
- link pataphysics to creativity - and computers
- ranking and pataphysics
-

7.1 Creativity

comparison table Poincare, Wallas, Polya, CC?

We had previously differentiated between creative computing and computational creativity:

> Intuitively the former is about doing computations in a creative way, while the latter is about achieving creativity through computation. You can think of the latter falling into the artificial intelligence

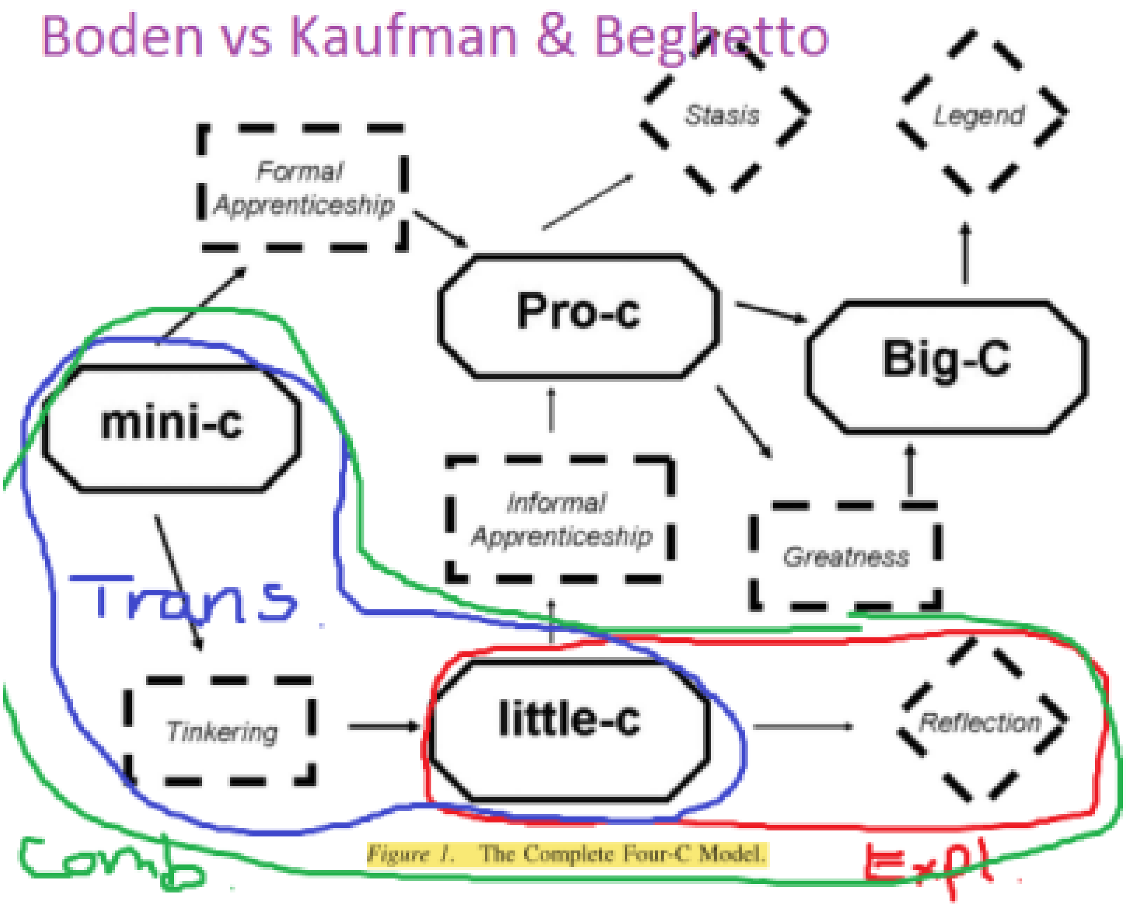

Figure 7.1: Kaufman's 4 C Model vs. Boden

> category (using formal computational methods to mimic creativity as a human trait, see also [18]) and the former being a more poetic endeavour of how the computing itself is done, no matter what the actual purpose of the program is. (Hugill 2013)

The differences are subtle but clear.

How does ethics fit into this?

7.2 Pataphysics

7.2.1 and Creativity

The creative process normally involves a move from the known to the unknown and sometimes from the named to the unnamed. In bringing something new

Creative Computing	**Digital Humanities**	**Computational Creativity**
Motivation	Design	Intentionality
Ideation	Curation, Analysis, Editing, Modeling and Networks, Infrastructure	Framing
Implementation	Computation, Processing	Process
Operation	Versioning, Prototyping, Failures	Product

Table 7.1: Comparison of Creative Computing vs Digital Humanities vs Computational Creativity

4 step model	**4 P Model**	**Problem Solving**
Preparation	Person	Understand
Incubation	Press/Context	Plan
Illumination	Process	Carry Out
Verification	Product	Look back

Table 7.2: Comparison of 4 Step Model vs 4 P Model vs Problem Solving

into existence, the human qualities of openness and tolerance of ambiguity are generally regarded as highly desirable. We may define creativity as **the ability to use original ideas to create something new and surprising of value**. We generally speak of creative 'ideas' rather than 'products', which merely provide evidence of a creative process that has already taken place. Both the originality and the value of an idea are evaluated using subjective criteria. **Pataphysics**, which represents an extreme form of subjectivity, is therefore a highly appropriate framework within which to encourage and enable creative thinking and operations.

Both the originality and the value of a creative idea are always evaluated using purely subjective criteria. Pataphysics, which represents an extreme form of subjectivity, is therefore a highly appropriate framework within which to encourage and enable creative thinking and operations.

> The ambiguity of experience is the hallmark of creativity, that is captured in the essence of pataphysics. Traversing the representations of this ambiguity using algorithms inspired by the syzygy, clinamen and anomaly of pataphysics, using a panalogical mechanism applied to metadata, should be able to humanize and even poeticize the experience of searching the Web. (Hendler and Hugill 2013)

Pataphysics is highly subjective and particular and is as such very suitable for this kind of transformation from relevant to creative.

> [Pataphysics] can only be defined in a new undiscovered language because too obvious: tautology. (Baudrillard 2007)

It is instructive to overlay these ideas on existing theories of creativity. Margaret Boden (Boden 2003), for example, has defined **P-creativity** (short for psychological creativity) as the personal kind of creativity that is novel in respect to the individual mind and **H-creativity** (short for historical creativity) as fundamentally novel in respect to the whole of human history. This allows for subjective evaluation of any idea.

see section baalh

Using Boden's definition we can call an idea 'new' if it is new to the individual who came up with it, making the idea P-creative. We can say that a creative idea can be seen from two perspectives: the subjective (P-creative) and the objective (H-creative) view. She argues that constraints support creativity, and are even essential for it to happen. "Constraints map out a territory of structural possibilities which can then be explored, and perhaps transformed to give another one" (Boden 2003, p.82).

This echoes the ideas of groups such as the **Oulipo** (which began as a Sub-Commission of the Collège de 'Pataphysique), who investigate 'potential literature' by creating constraints that frequently have a ludic element. Various other groups, the Ou-x-Pos, perform similar operations in fields as diverse as cinema, politics, music and cooking (Motte 2007).

Boden's conceptual space is the "territory of structural possibilities". So, the conceptual space of a teacup might be that it is meant to carry a certain amount of tea without breaking or burning fingers. It wouldn't be wise to create a teacup made out of paper. But whether we make a cup out of glass or porcelain, or how we shape the cup or the handle is pretty much up the individual's creativity. Being able to move around in this conceptual space, experiment (in thought or in reality) and play with different ideas while still following a given set of constraints is a good starting point for creativity to happen. Boden defines three sub-types of creativity.

The Oulipo similarly classifies its conceptual space under two broad headings: the synthetic and the analytic:

> [. . .] In the research which the Oulipo proposes to undertake, one may distinguish two principal tendencies, oriented respectively to-

> wards Analysis and Synthesis. The analytic tendency investigates works from the past in order to find possibilities that often exceed those their authors had anticipated. [. . .] The synthetic tendency is more ambitious: it constitutes the essential vocation of the Oulipo. It's a question of developing new possibilities unknown to our predecessors. This is the case, for example, of [Raymond Queneau's] 100,000,000,000,000 Poems or the Boolean haikus. (Motte 2007, p.27)

Later writings develop these ideas in more detail. La Littérature Potentielle **Oulipo1973**, is divided into several sections, dealing with clusters of methods, that include: anoulipisms (analytical oulipisms, such as combinatorial literature); use of preexisting structures such as lipograms (omitting a letter or letters), palindromes and snowballs (in which each successive word adds or subtracts a letter), homophonic translation, tautogram, and definitional literature; lexical, syntactic, or prosodic manipulations (such as the celebrated S+7, in which each substantive is replaced by the seventh word after it in a standard dictionary); lexicographical or prosodic synthoulipisms (early algorithmic methods); and perimathematical synthoulipisms (such as the Boolean poetry and combinatorial works already mentioned).

Boden links her three aspects of creativity to three sorts of surprise. She says that creative ideas are surprising because they go against our expectations. "The more expectations are disappointed, the more difficult it is to see the link between old and new." (Boden 2003, p.84) This suggests that fewer **expectations** (an open mind) allow creativity to happen more easily. Empirical experiences form expectations, which hinder our ability to accept creative ideas when they happen. In order to be able to recognise creative ideas we need to be able to see what they all have in common and in what way they differ and not reject unusual, unexpected ones.

> Unless someone realizes the structure which old and new spaces have in common, the new idea cannot be seen as the solution to the old problem. Without some appreciation of shared constraints, it cannot even be seen as the solution to a new problem intelligibly connected with the previous one. (Boden 2003, p.84)

It is clear that the Oulipo has a similar approach in its theorising of potential literature. Releasing creativity through constraint is its essential raison d'être.

This is not to say that experience and knowledge are necessarily bad for creativity. To appreciate creativity we need to be knowledgeable in the relevant domain to be able to recognise old and new connections and transformations. But we also need a certain level of openness and tolerance for ambiguity to overcome our expectations. Perhaps it is for this reason that 'creative people' are often assumed to have particular personality traits. Sternberg (Sternberg 1999,?), for example, proposes that these comprise: independence of judgement, self-confidence, and attraction to complexity, aesthetic orientation, and tolerance for ambiguity, openness to experience, psychoticism, risk taking, androgyny, perfectionism, persistence, resilience, and self-efficacy. More empirically, Heilman, Nadeau and Beversdorf (Heilman et al. 2003) have investigated the possible brain mechanisms involved in creative innovation. While a certain level of domain specific knowledge and special skills are necessary components of creativity, they point out that "co-activation and communication between regions of the brain that ordinarily are not strongly connected" might be equally important.

Newell, Shaw and Simon add to the above with their report on the creative thinking process (Newell et al. 1963). They identify three main conditions for creativity: the use of imagery in problem solving; the relation of unconventionality to creativity; and the role of hindsight in the discovery of new heuristics. Other issues they point out are abstraction and generalisation. So, for example, poets transform the grammar of their conceptual space (in this case, language) to create new sentence structures in a poetic form. By doing so, they go against the expectations, the possibilities of the language and cause surprise. Some people might not understand the transformations and therefore the jokes or beauty of a poem simply because they are either not able to recognise connections between the old and newly transformed elements (maybe due to a lack of knowledge in the poems topic or in that particular language) or because they do not want to accept unconventional methods.

Consider the following table, which compares some of the key ideas of creativity (Koestler 1964; Boden 2003; Indurkhya 1997) with the main pataphysical operations. It will be seen that pataphysics succeeds in bringing into sharp relief the more generalised scientific ideas. The pataphysical terms are taken from the natural sciences or philosophy, but always with an ironic twist, betraying their underlying humour. They connect quite strongly with the primary descriptors of creativity, while adding a certain layer of jouissance. **Pataphysics is self-avowedly useless, but its principles may prove surprisingly useful within this context.**

7.2.2 and Computers

CREATIVITY	**PATAPHYSICS**
Combinational: Juxtaposition of dissimilar, bisociation, deconceptualisation	**Antinomy**: Symmetry, duality, mutually incompatible, contradicting, simultaneous existence of mutually exclusive opposites **Syzygy**: Alignment of three celestial bodies in a straight line, pun, conjunction of things, something unexpected and surprising
Exploratory: Noticing new things in old places	**Anomaly**: Exceptions, equality
Transformative: Making new thoughts possible by transforming old conceptual space, altering its own rules	**Clinamen**: Unpredictable swerve, the smallest possible aberration that can make the greatest possible difference

Table 7.3: Creativity vs Pataphysics

ANDREW:
Since our solutions will be imaginary, our aim is not so much to have the computer generate creative artefacts as to engage in a creative dialogue with the user. Therefore, we do not intend to move as close to artificial intelligence as Colton's framework seems to suggest. In the pataphysical universe, ideas such as 'human skill', 'human imagination' and 'human appreciation' are too generalised to be useful. One may very well ask: which human? And when, where and even why? Rather, our project will aim to produce an exceptional computational entity that consistently generates surprising and novel provocations to the users, who in turn may navigate and modify these by deploying their own skills, appreciation and imagination. The relationship between the two will develop quite rapidly into one of mutual subversion since, however apparent the 'rules of the game' may become, the outcomes will always be particular or exceptional.

We are not the first people to attempt to apply pataphysical ideas in computer science. Johanna Drucker focused specifically on the cleft between formal logic and subjective judgement. She introduced the discipline of 'Speculative Computing' as a solution to that problem (Drucker and Nowviskie 2007). The concept can be understood as a criticism of mechanistic, logical approaches that distinguish between subject and object.

> Speculative computing takes seriously the destabilization of all cat-

> egories of entity, identity, object, subject, interactivity, process, or instrument. In short, it rejects mechanistic, instrumental, and formally logical approaches, replacing them with concepts of autopoiesis (contingent interdependency), quantum poetics and emergent systems, heteroglossia, indeterminacy and potentiality, intersubjectivity, and deformance. Digital Humanities is focused on texts, images, meanings, and means. Speculative Computing engages with interpretation and aesthetic provocation. (Drucker 2009, p.29)

For Drucker, aesthesis (ambiguous and subjective knowledge) is fundamentally opposed to mathesis (formal objective logic) and subjectivity is always in opposition to objectivity. Knowledge is a matter of interpretation of information, which can be represented digitally as data and metadata. She introduces what she calls a **'patacritical'** method of including exceptions as rules, even if repeatability and reliability are compromised. Bugs and glitches are privileged over functionality, and are "valuable to speculation in a substantive, not trivial, sense." As she says: "'Pataphysics inverts the scientific method, proceeding from and sustaining exceptions and unique cases" (Drucker and Nowviskie 2007).

In order to break out of the formal logic and defined parameters of computer science, she asserts, we need speculative capabilities and pataphysics. "The goal of pataphysical and speculative computing is to keep digital humanities from falling into mere technical application of standard practices." She links interface design with other speculative computing principles, and refers to Kant's idea of art as **'purposiveness without purpose'**. She says that the appreciation of design as a thing in itself (regardless of utility) is a goal of speculative aesthetics.

The projects Johanna Drucker describes in her book SpecLab (Drucker 2009) could certainly be considered related work. Not only in their theoretical foundations but also in some aspects of their implementation. One project in particular is worth mentioning here: the 'Patacritical Demon, an "interactive tool for exposing the structures that underlie our interpretations of text", although it remained a purely conceptual piece of work and was never implemented. Her idea if the "patacritical" method is quite interesting. Pataphysical exceptions and anomalies can thus be justified in a computational system. But it is not just this concept that deserves mention here. Her ideas on structured data, metadata and knowledge representation link very nicely into my project. How can we represent and structure data so that it does not lose its subjectivity, context and meaning? Her reference to graphical analogies is inspiring in that regard as well. I am certain I will refer back to her concepts throughout my thesis.

7.3 Patalgorithms

The proposed concept for a pataphysical algorithm requires precise data structures to represent the transformations that have taken place during the pataphysicalisation, such as the patadata. The system's index has to be adapted to accommodate this new type of data structure. It also needs to be flexible enough to allow algorithms to fit in at different stages or locations of the system, for example the inverted-index, ranking functions or query itself. Whilst this new style of algorithm has been proposed, current architectures are not capable of supporting them. As such a new flexible component-based software architecture has been proposed which will allow for a range of different style systems to be developed with little overhead. As such improving the chance of creative outcomes occurring in a different way.

7.3.1 Pataphylicalisation

The conceptual space for our project is 'pataphysical Web searching'. There are some very simple rules or constraints that form an initial definition of the project. For example it is clear that we want to search the World Wide Web (rather than a library database), that we want to return a list of search results (and not a pile of books) and that we want the search process and its results to be creative/pataphysical (rather than relevant). In a more technical sense, we have the query term(s), the index (of all web pages that we have crawled) and some pataphysical rules in our conceptual space. How we structure our search system, how we format the index or how we go about finding our results, is not in our conceptual space however. We can explore the space to its limits and we can transform it if we want to or feel like we need to. Our pataphysical rule set will include methods for transforming the space. By applying pataphysical rules to find results to our query we are pataphysicalising the query.

Definitions:

To pataphysicalise
: (verb) – applying pataphysical transformations

Pataphysicalisation
: (noun) – the process of pataphysicalising

Patadata
: (noun) – any data which has been pataphysicalised

But what exactly does the process of pataphysicalisation include? The kinds of transformations we are thinking of could be for example replacing or adding to the query term(s) with synonyms, antonyms, opposites, syzygies, clinamens etc. This can be done with the help of thesauri or dictionaries and ontologies. Whether we pataphysicalise our query term(s), the index or the results does not matter at this point. They are all possible and will maybe be done all at the same time (see figure 7.2 below). We can consider the possibility of a 'patametric index', rather than a parametric index or a 'patasaurus' (pataphysical thesaurus/ontology).

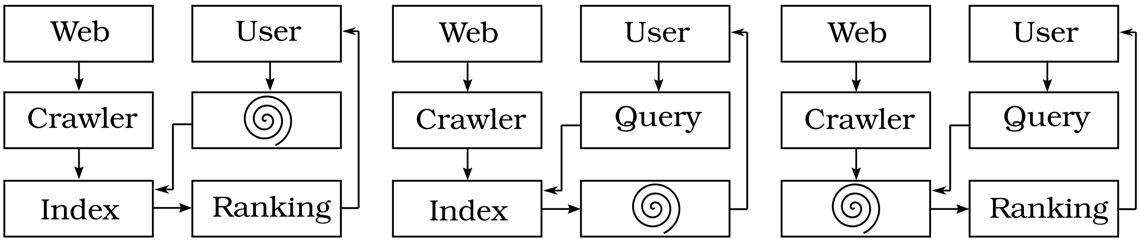

Figure 7.2: Pataphysicalisation

> Arguably, few other textual forms will have greater impact on the way we read, receive, search, access, use and engage with the primary materials of humanities studies than the metadata structures that organize and present that knowledge in digital form. (Drucker 2009, p.9)

7.3.2 Patadata

The idea of patadata is derived from the idea below:
Physics → Metaphysics → Pataphysics
Data → Metadata → Patadata

Patadata will allow us to engage with digital knowledge in a more creative way even. If metadata helps us organise information semantically then patadata is for organising information pataphysically. If metadata is objective then patadata is subjective. Drucker also points out that "many information structures have graphical analogies and can be understood as diagrams that organise the relations of elements within the whole." (Drucker 2009, p.16) So maybe patadata could allow us to represent these graphical analogies in some way? An alphabetical list is a typical model for representing text data sets for example. Or an otherwise ranked list, a tree structure, a matrix, a one-to-many relationship, etc. But is a ranked list really the best way to represent search results?

Ranking itself seems unpataphysical. It contradicts the philosophy of pataphysics, although we can argue that this contradiction makes it pataphysical again. Maybe this dilemma can be solved simply by adopting another type of graphical analogy to structure the results such as a tree structure instead of a ranked list.

Example: Let's say our patadata is represented by a list of keywords that each stands for a pataphysicalisation of the original query term. This list is added to each item in the index.

Query = 'Tree'
Patadata = [Tree(equivalent), Car(opposite), Paper(antinomy), Narwhal(anomaly), Book(syzygy), Venus Fly Trap(clinamen)]

Query = 'Sun God Ra'
Patadata = [Sun God Ra(equivalent), Slave(opposite), Holiday(antinomy), Blue Balloon(anomaly), Pyramid(syzygy), Sphinx(clinamen)]

7.3.3 Ranking - Pranking?

In traditional Web search, ranking signals 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 ideas such as the Facebook 'like' button or the Google '+1' button which could be seen as direct user relevance feedback.

Ranking can be done at different stages of the search process. Depending on how the index is formatted and what information can be pre-computed at that stage, the ranking algorithm evaluates every Web page for relevance and returns them in order. There exist lots of different approaches on ranking, including PageRank (Brin and Page 1998a) and HITS (Kleinberg 1999), which both analyse the link structure of the World Wide Web. They analyse the incoming and outgoing links on pages. PageRank for example assigns a numerical weight to each document, where each link counts as a vote of support in a sense. It is executed at indexing time, so the ranks are stored with each page directly in the index. 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 pages with many outgoing links are called hubs.

Given a query term X, what is considered a relevant match though? Do we simply return a list of Web pages where X appears in the heading of each page? It is obviously not that easy. Several ranking signals are combined together; Google states that they use over 200 signals including PageRank and they personalise results using signals such as the web history and location (Google n.d.). What kinds of ranking signals do we need for our pataphysical Web search tool? We could say that a page Y is relevant if it matches the patadata for query X. So, for example, Y would be a relevant result if it is a clinamen or syzygy to X. The more patadata matches there are the higher the ranking maybe. We don't necessarily have to assign a numerical ranking value to each page. Depending on how we structure our results page that might not be necessary. Shuffling the results list or the results tree could be an option.

Chapter 8

Practical Implementation

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

---

style inline code

run code on laptop and get snippets of all variable contents, e.g. faustroll, froll_dict, . . .

8.1 Setup

In TEXTSURFER.PY

Variables:

book_list
: List of books in library: `= PlaintextCorpusReader(corpus_root, '.*\.txt')`

troll
: Faustroll text split into words: `= book_list.words('faustroll.txt')`

faustroll
: Faustroll words converted to NLTK Text: `= nltk.Text(troll)`

faustroll_dict
: Alphabetical dictionary (list) of words: `= sorted(set([w for w in faustroll]))`

froll_dict
: faustroll_dict minus numbers and stopwords: `= [w for w in faustroll_dict`

```
if w.isalpha()not in sw]
```

8.2 Implementation

Interface (first tier) – application (second tier) – database/corpus (third tier)

The general concept of the project described in this paper is pataphysical web searching and the following three points summarize its main aims:

- search the Web for suitable answers to a given query,
- return results as a list or a mixture of data structures, and
- present pataphysical results (rather than relevant ones).

The essence of the proposed search tool lies in its algorithms which make the difference to traditional search engines. The philosophical ideology behind the tool is fundamentally different. Our system will still consist of the main components typically found in Web search engines (crawler, index and ranking) but they will have slightly different inner workings and target a different audience of users.

To link back to some of the creative, pataphysical concepts we have discussed earlier, let us put some of the ideas for our tool into perspective. The constraints for our conceptual space are the pataphysical rules that we want to apply to our data. We use those rules to explore, combine and transform our space; giving us the flexibility and freedom we need to find interesting results.

We developed the idea of pataphysicalising data as the process of applying such pataphysical rules in order to produce creative search results. This pataphysicalisation process forms a central component of our system (see Figure 8.2) and influences all areas of the search tool.

Our index will contain what Hendler and Hugill have called patadata [15]. Patadata is to metadata as metadata is to data - inspired by one of the definitions of pataphysics: that which is above that which is after physics [20]. This suggests that patadata provides another layer of information above information. If metadata helps us organise information semantically then patadata is for organising information pataphysically. If metadata is objective then patadata is subjective and that is precisely what pataphysics is for.

redraw figure

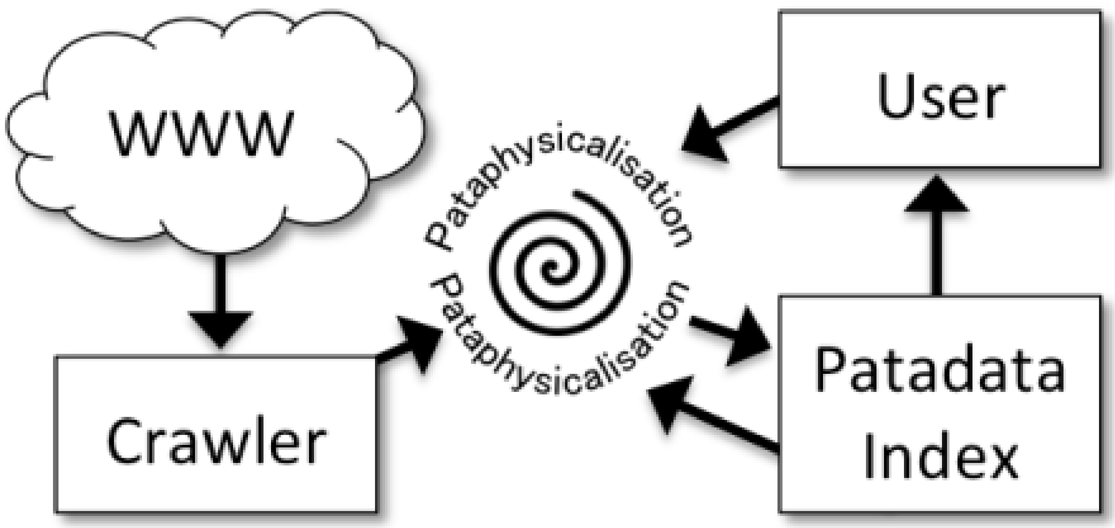

Figure 8.1: Pata central

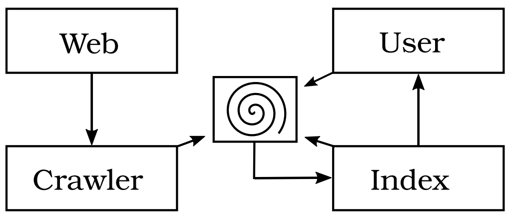

Figure 8.2: Pata central

The prototype described here (see Figure 2) was developed as a proof-of-concept tool to demonstrate some example search results using pataphysical algorithms.

It is by no means a fully functioning Web search engine, in fact it does not search the Web at all, and only the main algorithmic functionality is discussed here. The tool searches the text of an example book, namely Alfred Jarry's Exploits and Opinions of Dr. Faustroll, Pataphysician [20].

In short, the prototype's workflow can be described as follows:

- tokenise text and remove stopwords to build index,
- query triggers the three pataphysical functions,
- each function finds matches for query as described above,
- retrieve some words before/after match for context, and
- return list of resulting sentences.

The three functions inspired by pataphysics (clinamen, syzygy and antinomy) are described in more detail in the previous section. Figure 2 shows a screenshot of the resulting list of results for the query clear. The specific results for each of the three methods are simply a few words surrounding the pataphysicalised query term from within the book, which does not necessarily represent complete sentences but provides some context for the result.

The same principles and algorithms can be applied to different types of media, for example images or video and even sound. The complete tool would include a mixture of different types of media in its results with various styles of displaying them (e.g. displaying images ordered as a Fibonacci spiral).

8.2.1 Prototype 1

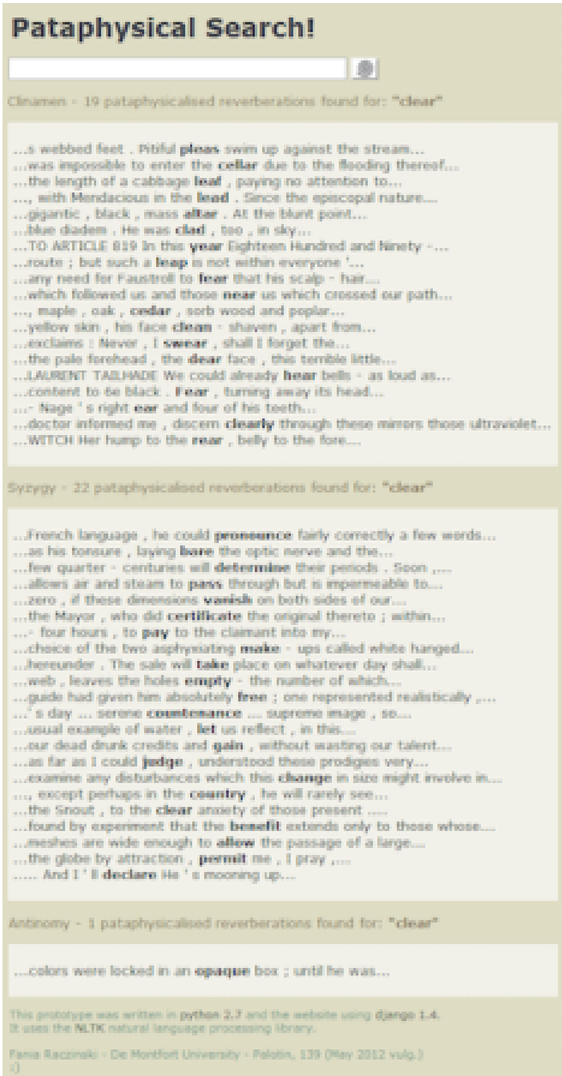

Figure 8.3: Prototype1

8.2.2 Prototype 2

Figure 8.4: Prototype2

```
# TEXT

## setup

1. read in faustroll book
2. create ‘froll_dict‘ dictionary from text
3. remove stopwords and numbers from ‘froll_dict‘

## text algorithm

1. get query term
2. execute three functions:

  2a. syzygy algorithm
```

```
    1. get list of synonyms
    2. for each synonym do the following:
      a. find hyponyms; if a hyponym occurs in ‘froll_dict‘ then add to the output lis
      b. find hypernyms; if a hypernym occurs in ‘froll_dict‘ then add to the output l
      c. find holonyms; if a holonym occurs in ‘froll_dict‘ then add to the output lis
    3. return list of syzygy words

  2b. antinomy algorithm
    1. get list of synonyms
    2. for each synonym do the following:
      a. find antonyms; if a antonym occurs in ‘froll_dict‘ then add to the output lis
    3. return list of antinomy words

  2c. clinamen algorithm
    1. find list of words within ‘froll_dict‘ that have a ‘dameraulevenshtein distance
    2. return list of clinamen words

3. get sentences for all three output lists

  3a. if the word appears in faustroll then find the nearest 5 words before and after
  3b. return list of sentences

4. render results as html


---

# IMAGES

## setup

- microsoft translate API key
- flickr API key
- (bing image search API key) - not used atm

## image algorithm

1. get query word
2. get one syzygy word using syzygy algorithm 2a above
3. translation party
  3a. translate english to french
```

```
  3b. translate french to japanese
  3c. translate japanese to english
4. get images
  4a. search flickr for 10 matches to english translation
  4b. get metadata for each
  4c. add title, thumb, link to output list
5. return output list
6. render results as html

---

# VIDEOS

## setup

- microsoft translate API key
- youtube API stuff
- (bing video search API key) - not used atm

## video algorithm

1. get query word
2. get one syzygy word using syzygy algorithm
3. translation party
  3a. translate english to french
  3b. translate french to japanese
  3c. translate japanese to english
4. get videos
  4a. search YouTube for 10 matches to english translation
  4b. get metadata for each
  4c. add title, thumb, link to output list
5. return output list
6. render as html
```

8.3 User experience

Whilst developing a system that returns creative results to the end user has numerous advantages, the assumptions that are made about and the decisions we take for the user must still be considered. For example, presume that the user inputs a search request 'The Cat in the Hat' after reading a Dr. Seuss book to their child, and the system employs an anomalous method on the query and searched 'sunglasses'. Whilst there is logic to the new search request, it is anomalous to the initial request, if the user receives these results without being told what method was used, the results will appear random, and therefore are likely to be detrimental to the user. Therefore the level of interaction the user has with the system and the feedback the system gives to the user on decisions it is making will have a large influence on the overall effectiveness and appreciation of the search tool. A quick and simple solution to this problem would be to add an icon to the side of each search result which displays how the original query was pataphysicalised.

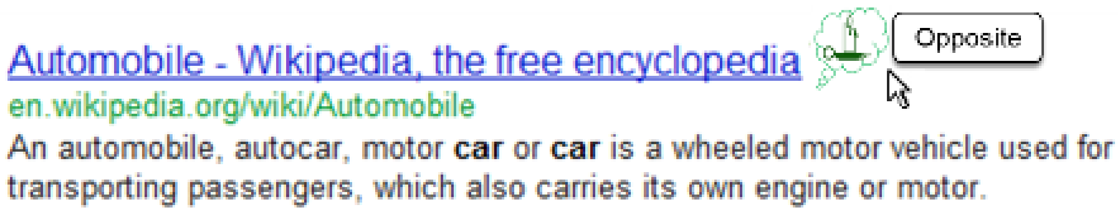

Figure 8.5: Feedback button

The above image (figure 3) shows an example of how this could be implemented. The little green candle (a reference to pataphysics in itself by the way) shows a pop-up note when hovered over with the mouse pointer. In this case the original query could have been 'tree' and 'car' was returned as an opposite to that.

In the end, it comes to a point of being able to identify which of these factors will affect how the user perceives the results and which do not, and therefore give the system greater flexibility. This in itself is a huge undertaking, with which large quantities of empirical data will be required and is therefore left for future work on the project.

8.4 Algorithms

8.4.1 Clinamen

The clinamen is the unpredictable swerve that Bök calls "the smallest possible aberration that can make the greatest possible difference" (Bök 2002).

The clinamen function uses the Damerau-Levenshtein algorithm [10], which measures the distance between two strings (with 0 indicating equality), to find

	clinamen	**syzygy**	**antinomy**
clear	altar, leaf, pleas, cellar	vanish, allow, bare, pronounce	opaque
solid	sound, valid, solar, slide	block, form, matter, crystal, powder	liquid, hollow
books	boot, bones, hooks, rocks, banks	dialogue, authority, record, fact	-
troll	grill, role, tell	wheel, roll, mouth, speak	-
live	love, lies, river, wave, size, bite	breathe, people, domicile, taste, see, be	recorded, dead

Table 8.1: Comparison of algorithms

words that are similar but not quite the same. The distance is calculated using insertion, deletion, substitution of a single character, or transposition of two adjacent characters. This means that we are basically forcing the program to return matches that are of distance two or one, meaning they have two or one spelling errors in them (see Eq. (1)). While we only return matches that actually appear in the book (i.e. they exist in the index), and by doing so eliminate the introduction of new words like Jarry's merdre, the swerve or aberration is still evident.

```
1 Clinamen(t):
2   matches = {v : 0 < dameraulevenshtein(t, v) <= 2}, for v $\in$ V
3   return matches
```

Line 1 of the algorithm above creates a set of matches, where each match is a word v from our vocabulary V if the Damerau-Levenshtein function computed with query term t returns a value less than or equal to 2 but higher than 0 (not the query term itself).

$$clinamen(t) = [v\colon 0 < dameraulevenshtein(t, v) \leq 2] \text{ for } v \in V \tag{8.1}$$

8.4.2 Syzygy

The syzygy surprises and confuses. It originally comes from astronomy and denotes the alignment of three celestial bodies in a straight line. In a pataphysical context it is the pun. It usually describes a conjunction of things, something unexpected and surprising. Unlike serendipity, a simple chance encounter, the syzygy has a more scientific purpose.

For the syzygy function, we made use of WordNet [29] through the NLTK python library [26] to find suitable results. Specifically, as shown in Eq. (2), the algorithm fetches the set of synonyms (synsets) first and then finds any hyponyms, hypernyms or holonyms for each of those (each of which denotes some sort of relationship or membership with its parent synonym). This mimics the syzygy alignment of three words in a line mentioned earlier (query $\rightarrow$ synonym $\rightarrow$ hypo-/hyper/holonym).

This algorithm makes heavy use of WordNet. Line 1 creates a set of all synonyms for query term t from WordNet. It then loops through all individual items in the list of synonyms in line 2 to 5. Line 3 adds all hyponyms h for the current synonym s if there exists a word in the vocabulary same as the hyponym. Similarly lines 4 and 5 add all hypernyms and holonyms to the results list which is then returned in line 6.

```
1 Syzygy( t ):
2   synonyms = { s $\in$ WordNet-synsets( t ) }
3   for each s in synonyms:
4     hypo  = { h : h $\in$ WordNet-hyponyms( s ) }
5     hyper = { h : h $\in$ WordNet-hypernyms( s ) }
6     holo   = { h : h $\in$ WordNet-holonyms( s ) }
7     union = hypo $\cup$ hyper $\cup$ holo
8     add { h : h $\in$ union and $\exists$ h $\in$ V } to results
9   return results
```

$$\begin{aligned}
syzygy(t) &= \{h\colon h \in union(t) \text{ and } \exists\, h \in V\} \\
union(t) &= hypo(t) \cup hyper(t) \cup holo(t) \\
hypo(t) &= \{h\colon h \in hyponyms(s)\}, \text{for } s \in syno(t) \\
hyper(t) &= \{h\colon h \in hypernyms(s)\}, \text{for } s \in syno(t) \\
holo(t) &= \{h\colon h \in holonyms(s)\}, \text{for } s \in syno(t) \\
syno(t) &= \{s\colon s \in synonyms(t)\}
\end{aligned} \tag{8.2}$$

8.4.3 Antinomy

The antimony, in a pataphysical sense, is the mutually incompatible.

For the antinomy we simply used WordNet's antonyms (opposites) (see Eq. (3)). This function exhibits the same problem as mentioned above for the syzygy, just much worse. Arguably, some words just do not appear to have an opposite, but the pataphysical antinomy should still be able to find a match. A better thesaurus or a larger index (e.g. based on more than one book – or, of course, the Web) could improve this method.

This algorithm is very similar to the algorithm for the syzygy. It finds all antonyms through WordNet and returns them if they appear in our dictionary of words.

```
Antinomy(t):
  synonyms = { s $\in$ WordNet-synsets(t) }
  for each s in synonyms:
    add { h $\in$ WordNet-antonyms(s) : $\exists$ h $\in$ V } to results
  return results
```

$$
\begin{aligned}
antinomy(t) &= \{h \colon h \in anto(t) \text{ and } \exists\, h \in V\} \\
anto(t) &= \{h \colon h \in antonyms(s)\}, \text{for } s \in syno(t) \\
syno(t) &= \{s \colon s \in synonyms(t)\}
\end{aligned} \tag{8.3}
$$

8.5 Website

http://pata.fania.eu

8.6 APIs

- Bing
- YouTube
- Microsoft Translate
- Flickr
- WordNet

8.7 Design

- random sentences
- spiral
- Word Scrable
- ...

Chapter 9

Impact and Applications - Case Study

In this section we consider the possible uses and applications for the proposed creative search tool.

Our target audience is not quite as broad as that of a general search engine like Google. Instead, we aim to specifically cater for users who can appreciate creativity or users in need of creative inspiration. Users should generally be educated about the purpose of the search tool so that are not discouraged by what might appear to be nonsensical results. Users could include artists, writers or poets but equally anybody who is looking for out-of-the-box inspirations or simply a refreshingly different search engine to the standard. The way we display and label results produced by the tool can influence how the user perceives them. The current prototype for example separates the results into its three components but we could have equally just mixed them all together. The less transparent the processes in the background (e.g. which algorithm was used, how does the result relate to the query precisely, etc.) are for the user, the more difficult it might be to appreciate the search.

There are many ways a pataphysical search tool could be used across disciplines. In literature, for example, it could be used to write or generate poetry, either practically or as a simple aid for inspiration. We are not limited to poetry either; novels, librettos or plays could benefit from such pataphysicalised inspirations. One can imagine tools using this technology that let you explore books in a different ordering of sentences (a sort of pataphysical journey of paragraph hopping), tools that re-write poems or mix and match them together. Even our simple prototype shows potential in this area and could be even more powerful if we extended it to include more base texts, for example the whole set of books contained in Faustroll's library ([20] and also [12]). A richer body of texts (by different authors) would produce a larger index which would possibly find many more matches through WordNet and end in a more varied list of results.

From a computer science perspective it could be used as one of the many algorithms used by traditional search engines for purposes like query feedback or expansion (e.g. "did you mean ... "or "you might also be interested in ... "). Depending on how creative we want the search engine to be, the higher we would rank the importance of this particular algorithm. One of the concepts related to the search tool, namely patadata, could have an impact on the development of the Semantic Web. Just as the Semantic Web is about organizing information semantically through objective metadata, patadata could be used to organize information pataphysically in a subjective way.

The prototype tool is already being used in the creation of an online opera, provisionally entitled from [place] to [place], created in collaboration with The Opera Group , an award-winning, nationally and internationally renowned opera company, specialising in commissioning and producing new operas. In particular, it is being used to create the libretto for one of the virtual islands whose navigation provides the central storyline for the opera. The opera will premiere in 2013, and will continue to develop thereafter, deploying new versions of the tool as they appear.

9.1 Patakosmos

www.patakosmos.com

9.2 Soeren and the other guy

9.3 Digital Opera

9.3.1 Use

> "There is an official and an unofficial way that I used the prototype. Officially, I threw keywords based on mood 'sad', 'lively' etc into it and used the results as the libretto for small sections of music

that reflect said mood. Unofficially I used lots and lots of different words to retrieve the lines that worked." Lee Scott (22 May 2014)

9.3.2 Result

Confusing
...my tuning fork. imagine the perplexity of a man outside time...
...mandrills or clowns, spread their caudal fins out wide like acrobats...
...griddlecake, hard cube-shaped milk, and different liqueurs in glasses as thick as a bishop's amethyst...

Playful
...peacocks' tails, gave us a display of dancing on the glassy...

Busy
...wasps and bumblebees and the vibration of a fly's wing...

Driving
...bodies striking the hours of union and division of the black...

Disjointed
...tangential point of the universe, distorting it according to the sphere's...

Sadness
...others: may your dire sorrow flyaway...
...no longer deep enough to satisfy our honour...
...other side of the green sleep of hulls; ships passed away...

Sweeping
...loved her like the infinite series of numbers...
...the veritable portrait of three persons of god in three escutcheons...

Fear
...it will set. fear creates silence nothing is terrifying...
...forth revealing the distinction and evil engraved in the wood...
... underground arose from ali baba screaming in the pitiless oil...

Joy
...sibyls record the formula of happiness, which is double: be amorous...
...the lord of the island gloried that his creation was good...

Awe
...like earth; the enemy of fire and renascent from it...

…awesome figure, warlike and sacerdotal, glared at the assembly…
…is not an island but a man…

Clocked
…quincuncial trees…

Tension
…the vigilant gaze of the spirit of the dead…
…do not make as much noise as a single drum…
… the oars made a clangourous sound as they scraped along the bow….

Calm
…a strange upon a clam sea quilted with sand; faustroll…
…each person present threw a pebble into the sea…
…depth and with edges that tend to ebb and flow…
(textures)

Morphing
…in a striking metamorphosis the mourning color of the hangings turned…

9.3.3 Interview

Part IV

MΣT∀-LΘGIC∀LYSIS

Chapter 10

Interpretation

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

Creativity should be seen as a continuum, there is no clear cut-off point or Boolean answer to say precisely when a person or piece of software has become creative or not.

Evaluating creative software is not an easy task. Pease and Colton [27] divide it into two notions:

- whether an idea or artefact is valuable or not, and
- whether a system is acting creatively or not.

We would need to investigate each individual search result in terms of its value and creativity. This could be done by user ratings or satisfaction questionnaires. Rather than measuring the success of individual results we could look at evaluation them as one set instead, similar to the blind side-by-side comparisons by Bing or MillionShort .

The search results produced by our tool can be quite surprising sometimes and it not always clear how they connect to the initial query (especially if the inner workings of the algorithms are unknown), even if we identify through which function a result has been obtained. Obviously these keywords might not be helpful to users unfamiliar with the concept of pataphysics and might therefore appear rather nonsensical. Whilst there is a clear logic to each search result, they might appear anomalous to the user's expectations if he received these results without knowing the philosophy of the search tool. The results could possibly appear random then, and would therefore likely to be detrimental to the user.

To prevent that, the level of interaction the user has with the system and the feedback the system gives to the user on decisions it is making will have a large influence on the overall effectiveness and appreciation of the tool. A quick and simple solution to this problem would be to add an icon to the side of each search result, which displays exactly how the original query was pataphysicalised.

Creativity does not have a universally accepted definition. Creativity is a human quality and definitions don't necessarily lend themselves to be applied to computers as well. There are aspects that come up in many, like novelty and value, but some that rarely pop up, like relevance and variety. Creativity can be studied at various **levels** (neurological, cognitive, and holistic/systemic), from different **perspectives** (subjective and objective) and **characteristics** (combinational, exploratory and transformative). Creativity should be seen as a continuum, there is no clear cut-off point or Boolean answer to say precisely when a person or piece of software has become creative or not.

Current evaluation methodologies have concentrated on only a handful of the points raised above, for example studying only the creative end-product itself (out of context), only judging it by its novelty objectively, assigning an arbitrary thresholds, etc. This also includes the assumption that machines "only" mimic humans and are therefore not judged by full potential.

What does it mean, how can it be measured?

Subjectivity vs objectivity is a theme throughout

How is it defined and measured in humans, what can we just take directly from those concepts and apply them directly to machines and what needs to be completely redefined?

This paper discusses issues related to the study of creativity in a computer science context. Two transdisciplinary fields of study have emerged from the variety of disciplines concerned. These are computational creativity and creative computing. The former lies at the cross section of artificial intelligence and cognitive science and the latter is mostly distinguished by its involvement in art. Creative computing focuses on the process of creativity rather than just the outcome as in computational creativity. In the remainder of this paper, CC will always denote creative computing unless otherwise specified.

Many of these (if not all) spawn from the computational creativity discipline.

Introduce the difference between human and computer evaluation/creativity?

> In research communities, approaches to the study of creativity differ in three main respects: 1) the type of research design, whether experimental, psychometric, observational etc. 2) the focus of the research, whether on human attributes cognitive processes or features of creative outcomes, and 3) the type of information that is used for the basis of evidence, by which is meant whether the time frame is present (real-time observation) or past (historical data) and whether the situation is artificial (laboratory) or natural (real world settings). (Candy 2012, p.3)

distinguishing between person's and product's creativity (Piffer 2012, p.258) it is concluded that a person's creativity can only be assessed indirectly (for example with self report questionnaires or official external recognition) but it cannot be measured (Piffer 2012, p.258)

10.0.4 Measurable Attributes

Novelty	**Value**	**Quality**	**Ephemeral[redacted] Uncontrolled**	**l Controlled**	**Purpose**
Originailty	Usefulness	Skill	Serendipity[redacted]	ce[redacted]	n
Newness	Appropriateness	Eff[redacted]	Randomness[redacted]		Communication[redacted]
Variety	Appreciation[redacted]	ce[redacted]	ty[redacted]	ent[redacted]	n
Typicality	Relevance	Intellect		Progression[redacted]	
Imagination[redacted]		Acceptability[redacted]		Spontaneity[redacted]	
	Influence		Experimentation[redacted]	ndependence[redacted]	

Table 10.1: Summary of all creativity attributes

Temporal, spatial, ephemeral. . . what else?

Summary
- Mimicking
- novelty + value + quality
- randomness + serendipity

10.1 Problems with Evaluation

Evaluating **human creativity** is problematic.

There are many debates across the involved disciplines. Specifically, Mayer identified five big questions of human creativity research: (Mayer 1999, p.450-451)

1. Is creativity a property of people, products, or processes?
2. Is creativity a personal or social phenomenon?
3. Is creativity common or rare?
4. Is creativity domain-general or domain-specific?
5. Is creativity quantitative or qualitative?

> An important challenge for the next 50 years of creativity research is to develop a clearer definition of creativity and to use a combination of research methodologies that will move the field from speculation to specification. (Mayer 1999, p.459)

Taking these debates about human creativity and directly applying them to machines seems logical but may be the wrong and lazy approach. Adapting Mayer's five big questions to machines does not seem to capture the real issues at play.

1. Is creativity a property of programmers, users, machines, products, or processes?
2. Is creativity a local, a network or an Internet phenomenon?
3. Is creativity common or rare? (P or H creativity)
4. Is creativity domain-general or domain-specific?
5. Is creativity quantitative or qualitative?

- Can a machine judge whether a human is creative?
- Is creativity a property of machines (hardware or software?)

- Is mimicking human creativity really enough and appropriate?
- Compare against "human creativity"? Or define machine creativity from scratch?
- Who is creative? The programmer or the program?
- Can creativity be objectively measured?
- Quantitative or qualitative?
- In respect to P or H creativity?
- Output minus input? (we don't have the same strict judgement on humans)
- Is it the product or the process or both?
- Does context matter? (Blind deaf dumb person = computer?)
- Does time matter?
- Does purpose or intention matter?
- AGI vs AI? Artificial general creativity vs artificial creativity?

On a more practical level, there are various problems that arise when trying to evaluate computer creativity. Anna Jordanous found that "evaluation of computational creativity is not being performed in a systematic or standard way" (Jordanous 2011, p.2)(her emphasis).

(neurological, cognitive and systemic) in the computer sense!

Since most problems with evaluating creativity in computers (and humans alike) stems from the lack of a universal definition it seems logical to try and remedy this first and foremost.

Creativity is studied in many disciplines.

- understanding the chemical mechanisms within the brain (neurological)
- understanding the thought processes associated with creativity (cognitive)
- understanding creativity in children and adults, novices and professionals
- understanding creativity in individuals and society (holistic)

Creativity is a continuum, which means that being creative and not being creative form the two distinct extreme ends of a scale.

> "a continuous sequence in which adjacent elements are not perceptibly different from each other, but the extremes are quite distinct" [OED]

(subjective and objective)

How can we model Koestler's bisociative creativity in computers? Boden/Kaufman: Subjective and objective types (pandh or little-candbig-C) (Boden 2003; Kaufman and Beghetto 2009) (product+process)

DIGITAL CREATIVITY ?!?!?! Mix between digital humanities and creative computing/computational creativity —- see Digital Creativity Journal!!!! Unified theory of creativity! (Koestler?) Unified definition!

1. What is the definition of creativity?
2. Who is being creative? WHO
3. What was the aim/intention, if any?
4. What was the process, approach? HOW
5. What factors influenced the person/process? WHERE
6. What is the result/product, if any? Is it original, relevant? WHAT
7. What is the impact, if any?
8. What is the maintenance plan, if any?

10.2 5 P's: product, process, people, place and purpose

> One way of characterizing these processes is to use an alliteration that allows us to keep track of some of the core features of RRI in ICT, namely the four P's, which are: product, process, purpose and people. The purpose of using the four P's is to draw attention to the fact that, in addition to the widely recog- nized importance of both product and process of technical development, the purpose of the development needs to be considered and people involved in the innovation need to be incorporated in RRI. (Stahl et al. 2013, p.203, my emphasis)

combine the 4 P's with purpose// 5 P's: product, process, people, place and purpose// Why is the purpose important?// Interpreting or Measuring?// Maybe we should not be looking for metrics but rather guidelines for interpretations of creativity.

In the end I believe it is impossible to measure creativity objectively. I don't just think it is impossible, I think it is unwise to try and do so. It would be silly to put a percentage on how creative something is just like it would be silly to say a certain product is 50percent ethical. In fact there are lots of parallels between (computer) ethics and (computer) creativity. Both are subjective, both are highly dependent on context.

What is important is to study and consider the factors that influence our perception of whether something is creative (or ethical) and what the implications are.

Creativity in a process or product will mean different things to different people, in different environments and contexts. Common sense.

Just as there are people who just cannot see any creativity in in modern art, there will always be people who wont accept anything produced by a computer as creative.

Chapter 11

Patacritical Analysis

11.1 Problems Encountered

11.2 Design Aspects

11.3 Search Results

Part V

H∀PPILY ΣVΣR ∀FTΣR ?

Chapter 12

Aspirations

PROBLEMS ENCOUNTERED AND SUGGESTED SOLUTIONS

SHORTCOMINGS AND MISSING FUNCTIONALITY

- Research in science and art
- Review paper? Pataphysics and creativity?
- Quantitative research questions
- Working definition for Pataphysics
- Examples for Pataphysics conepts
- Examples for types of creativity
- Examples for creative process
- Explain learys tables
- How can we use creative concepts discussed?

12.1 Code

FURTHER DEBUGGING OF CODE (IF NECESSARY)

12.2 Interface

DESIGN ASPECTS IMPROVEMENTS/ALTERNATIVES TO USER INTERFACE DESIGN

12.3 Algorithms

IMPROVEMENTS/ALTERNATIVES TO ALGORITHMS

12.4 Architecture

IMPROVEMENTS/ALTERNATIVES TO ARCHITECTURE

12.5 Research

USER FEEDBACK (IF NECESSARY)

Chapter 13

Observations

From here, we can try to implement different algorithms or different pataphysical concepts within our existing tool or built a different system. The next logical step would be to implement a fully functioning Web search engine using the algorithms described in this paper. But before we go into further development, it might be worth evaluating and interpreting the results produced by the prototype.

In this paper we have introduced a new approach for a creative search tool that uses pataphysics as an underlying philosophy. We have explained how pataphysics can be used in search algorithms to produce interesting results with a humorous twist. Our initial experiments within a limited domain have shown that the generated results can indeed be novel, surprising and useful. We have also briefly discussed ideas for applications of the tool and issues that may trigger possible further research in in the field of Computing. We have presented some thoughts on evaluation of our tool and future work.

13.1 Achievements

13.2 Implications

13.3 Recommendations

POST☹

Appendix A

Code

setup
text setup
video setup
image setup

Clinamen
algorithm
get results
render results

etc

A.1 Setup

```
from __future__ import print_function
from flask import url_for
from nltk.corpus import wordnet as wn
from nltk.corpus import PlaintextCorpusReader
import os
import nltk
import codecs
import sys
import warnings
root_path = os.path.dirname(os.path.abspath(__file__))
root_path = root_path[:-4]
corpus_root = root_path + '/app/static/corpus'
book_list = PlaintextCorpusReader(corpus_root, '.*\.txt')
troll = book_list.words('faustroll.txt')
```

```
faustroll = nltk.Text(troll)
path_e = corpus_root + '/english'
stopwords_doc = open(path_e, "r")
sw = [i for line in stopwords_doc.readlines() for i in line.split()]
stopwords_doc.close()
faustroll_dict = sorted(set([w for w in faustroll]))
froll_dict = [w for w in faustroll_dict if w.isalpha() not in sw]
```

A.2 Clinamen

```
def clinamen(word, i):
    out = set()
    items = [item for item in froll_dict if dameraulevenshtein(word, item) <= i]
    for item in items:
        if item != word:
            out.add(item)
    return out
```

A.3 Damerau Levenshtein

```
def dameraulevenshtein(seq1, seq2):
    oneago = None
    thisrow = range(1, len(seq2) + 1) + [0]
    for x in xrange(len(seq1)):
        # Python lists wrap around for negative indices, so put the
        # leftmost column at the *end* of the list. This matches with
        # the zero-indexed strings and saves extra calculation.
        twoago, oneago, thisrow = oneago, thisrow, [0] * len(seq2) + [x + 1]
        for y in xrange(len(seq2)):
            delcost = oneago[y] + 1
            addcost = thisrow[y - 1] + 1
            subcost = oneago[y - 1] + (seq1[x] != seq2[y])
            thisrow[y] = min(delcost, addcost, subcost)
            # This block deals with transpositions
            if (x > 0 and y > 0 and seq1[x] == seq2[y - 1]
                    and seq1[x - 1] == seq2[y] and seq1[x] != seq2[y]):
                thisrow[y] = min(thisrow[y], twoago[y - 2] + 1)
    return thisrow[len(seq2) - 1]
```

A.4 Antinomy

```
def antinomy(word):
```

```
    out = set()
    wordsets = wn.synsets(word)
    for w in wordsets:
        anti = w.lemmas()[0].antonyms()
        if len(anti) > 0:
            for a in anti:
                if str(a.name()) != word:
                    out.add(str(a.name()))
    return out
```

A.5 Syzygy

```
def syzygy(word):
    out = set()
    wordsets = wn.synsets(word) # returns a list of synsets
    for w in wordsets: # w is a synset
        # Hyponyms share a type-of relationship with their hypernym
        hypo = w.hyponyms() # returns a list of synsets
        if len(hypo) > 0:
            for h in hypo: # h is a synset
                for l in h.lemmas(): # l is a lemma
                    if str(l.name()) in froll_dict:
                        out.add(str(l.name()))
        # Hyponyms share a type-of relationship with their hypernym
        hyper = w.hypernyms()
        if len(hyper) > 0:
            for h in hyper:
                for l in h.lemmas():
                    if str(l.name()) in froll_dict:
                        out.add(str(l.name()))
        # 'X' is a holonym of 'Y' if Ys are parts of Xs, or
        # 'X' is a holonym of 'Y' if Ys are members of Xs.
        holo = w.member_holonyms()
        if len(holo) > 0:
            for h in holo:
                for l in h.lemmas():
                    if str(l.name()) in froll_dict:
                        out.add(str(l.name()))
    return out
```

A.6 Random Quotes

```
def getrandquote():
  # RANDOM QUOTES
```

```
    root_path = os.path.dirname(os.path.abspath(__file__))
    root_path = root_path[:-4]
    corpus_root = root_path + '/app/static/corpus'
    path_b = corpus_root + '/quotes.txt'
    quotes_text = codecs.open(path_b, "r", encoding='utf-8')
    quotestext = quotes_text.readlines()
    quotes_text.close()
    # print random.choice(quotestext)
    return random.choice(quotestext)
```

A.7 Text Results

```
from flask import render_template, url_for, request
from app import app
from textsurfer import *


@app.route('/text')
def text():
    return render_template('text.html')


@app.route('/textresults', methods=['GET', 'POST'])
def textresults():

    query = request.form['query']

    if request.method == 'GET':
        print 'textresults get: ', query # data['query']
        # return render_template('p01results.html', q)
    else:
        #request was a POST
        print 'textresults post: ', query # data['query']
        # qx = getResults(q)

       # CLINAMEN
        sens = dict([])
        pre_sens = dict([])
        post_sens = dict([])
        clinamen_words = clinamen(query, 2)
        clinamen_len = len(clinamen_words)
        for r in clinamen_words:
            if len(pre_sentence(r)) > 0:
                pre_sens[r] = pre_sentence(r)
            if len(post_sentence(r)) > 0:
```

```
                post_sens[r] = post_sentence(r)
            if len(find_sentence(r)) > 0:
                sens[r] = find_sentence(r)
        # SYZYGY
        syssens = dict([])
        pre_syssens = dict([])
        post_syssens = dict([])
        syzygy_words = syzygy(query)
        syzygy_len = len(syzygy_words)
        for r in syzygy_words:
            if len(pre_sentence(r)) > 0:
                pre_syssens[r] = pre_sentence(r)
            if len(post_sentence(r)) > 0:
                post_syssens[r] = post_sentence(r)
            if len(find_sentence(r)) > 0:
                syssens[r] = find_sentence(r)
        # ANTINOMY
        antisens = dict([])
        pre_antisens = dict([])
        post_antisens = dict([])
        antinomy_words = antinomy(query)
        antinomy_len = len(antinomy_words)
        for r in antinomy_words:
            if len(pre_sentence(r)) > 0:
                pre_antisens[r] = pre_sentence(r)
            if len(post_sentence(r)) > 0:
                post_antisens[r] = post_sentence(r)
            if len(find_sentence(r)) > 0:
                antisens[r] = find_sentence(r)
                #print antisens[r]

        #print data
        return render_template('textresults.html', **locals())
```

A.8 Pataphysicalise

```
def pataphysicalise(word):

    translator = Translator(microsoft_id, microsoft_secret)
    syzygy_words = textsurfer.syzygy(word)
    if syzygy_words:
        syzword = syzygy_words.pop()
    else:
        syzword = word
    # Pataphysicalisation !!
```

```
    french = translator.translate(syzword, "fr")
    japanese = translator.translate(french, "ja")
    patawords = translator.translate(japanese, "en")
    translations = (french, japanese, patawords)
    return translations
```

A.9 Getting Videos

```
def getvideos(word):
    out = []

    translations = pataphysicalise(word)
    # translations = word
    patawords = translations[2]

    # YOUTUBE
    client = yt.YouTubeService()
    query = yt.YouTubeVideoQuery()
    query.vq = patawords
    query.max_results = 10
    query.start_index = 1
    query.racy = 'exclude'
    query.orderby = 'relevance'
    query.safeSearch = 'strict'
    query.format = '5'

    feed = client.YouTubeQuery(query)

    for entry in feed.entry:
        title = entry.media.title.text
        tags = entry.media.keywords.text
        author = entry.author
        url = entry.GetSwfUrl()
        thumb = entry.media.thumbnail[0].url
        out.append((title, thumb, url))

    # CHECK CODE BELOW!!!!
    ## BING VIDEOS
    #params =
        "Video?$format=json&VideoFilters='Aspect:Square'&$top=5&Query='%s'" %
        #patawords
    ## &ImageFilters='Size:Small+Aspect:Square'
    #url = base + params
    #bing_img = requests.get(url, auth=HTTPBasicAuth(None, key))
    #for result in bing_img.json['d']['results']:
```

```
    #   phototitle = ud.normalize('NFKD', result['Title']).encode('ascii',
        'ignore')
    #   photothumb = result['Thumbnail']['MediaUrl']
    #   photolink = result['SourceUrl']
    #   out.append((phototitle, photothumb, photolink))
    return out, translations
```

A.10 Getting Images

```
def getvideos(word):
    out = []

    translations = pataphysicalise(word)
    # translations = word
    patawords = translations[2]

    # YOUTUBE
    client = yt.YouTubeService()
    query = yt.YouTubeVideoQuery()
    query.vq = patawords
    query.max_results = 10
    query.start_index = 1
    query.racy = 'exclude'
    query.orderby = 'relevance'
    query.safeSearch = 'strict'
    query.format = '5'

    feed = client.YouTubeQuery(query)

    for entry in feed.entry:
        title = entry.media.title.text
        tags = entry.media.keywords.text
        author = entry.author
        url = entry.GetSwfUrl()
        thumb = entry.media.thumbnail[0].url
        out.append((title, thumb, url))

    # CHECK CODE BELOW!!!!
    ## BING VIDEOS
    #params =
        "Video?$format=json&VideoFilters='Aspect:Square'&$top=5&Query='%s'" %
        #patawords
    ## &ImageFilters='Size:Small+Aspect:Square'
    #url = base + params
    #bing_img = requests.get(url, auth=HTTPBasicAuth(None, key))
```

```
    #for result in bing_img.json['d']['results']:
    #    phototitle = ud.normalize('NFKD', result['Title']).encode('ascii', 'ignore')
    #    photothumb = result['Thumbnail']['MediaUrl']
    #    photolink = result['SourceUrl']
    #    out.append((phototitle, photothumb, photolink))
    return out, translations
```

Bibliography

Jorge Luis Borges. John Wilkins' Analytical Language. In Eliot Weinberger, editor, *Selected Non-Fictions*, pages 229–232. Penguin Books, London, 2000. URL http://www.entish.org/essays/Wilkins.html.

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

Jim Hendler and Andrew Hugill. The syzygy surfer: (Ab)using the semantic web to inspire creativity. *International Journal of Creative Computing*, 1(1):20–34, 2013.

Alfred Jarry. *Exploits and Opinions of Dr Faustroll, Pataphysician*. Exact Change, Cambridge, MA, 1996. ISBN 1878972073.

Push Singh. *EM-ONE: An Architecture for Reflective Commonsense Thinking*. PhD thesis, Massachusetts Institute of Technology, 2005. URL http://web.media.mit.edu/~push/push-thesis.html.

Google. Google Ranking, 2012. URL https://www.google.com/intl/en/about/company/philosophy/.

Johanna Drucker. *SpecLab: Digital Aesthetics and Projects in Speculative Computing*. University of Chicago Press, 2009. ISBN 0226165086. URL http://books.google.com/books?id=VPXCk396uPYC&pgis=1.

Christian Bök. *'Pataphysics: The Poetics of an Imaginary Science*. Northwestern University Press, Evanston, Illinois, 2002. ISBN 0810118777. URL http://books.google.com/books?id=Dzd7LyYblf8C&pgis=1.

Andrew Hugill. *'Pataphysics: A Useless Guide*. MIT Press, Cambridge, Massachusetts, 2012a. ISBN 9780262017794.

Fania Raczinski, Hongji Yang, and Andrew Hugill. Creative Search Using Pataphysics. In *Proceedings of the 9th International Conference on Creativity and Cognition*, pages 274–280, Sydney, Australia, 2013. ACM New York, NY, USA. ISBN 978-1-4503-2150-1.

James Sawle, Fania Raczinski, and Hongji Yang. A Framework for Creativity in Search Results. In *The Third International Conference on Creative Content Technologies*, pages 54–57, Rome, 2011.

Andrew Hugill, Hongji Yang, Fania Raczinski, and James Sawle. The pataphysics of creativity: developing a tool for creative search. *Digital Creativity*, 24(3):237–251, September 2013. ISSN 1462-6268. doi: 10.1080/14626268.2013.813377. URL http://www.tandfonline.com/doi/abs/10.1080/14626268.2013.813377.

V. Ramesh, Robert L. Glass, and Iris Vessey. Research in computer science: an empirical study. *Journal of Systems and Software*, 70(1-2):165–176, February 2004. ISSN 01641212. doi: 10.1016/S0164-1212(03)00015-3. URL http://linkinghub.elsevier.com/retrieve/pii/S0164121203000153.

Basarab Nicolescu. Methodology of Transdisciplinarity – Levels of Reality, Logic of the Included. *Transcdisciplinary Journal of Engineering and Science*, 1(1): 19–38, 2010.

Alastair Brotchie, Stanley Chapman, Thieri Foulc, and Kevin Jackson, editors. *'Pataphysics: Definitions and Citations*. Atlas Press, London, 2003. ISBN 1900565080. URL http://www.atlaspress.co.uk/theLIP/index.cgi?action=departmental.

Andrew Hugill. *Lineaments of 'Pataphysics*. Self, 2012b.

Alastair Brotchie. *A supplement*. Atlas Press, UK, 2011. ISBN 978-1-900565-55-4.

Dave Walsh. Absinthe, Bicycles and Merdre, 2001. URL http://old.disinfo.com/archive/pages/dossier/id975/pg1/index.htmlhttp://www.blather.net/blather/2001/05/alfred_jarry_absinthe_bicycles.html.

Alastair Brotchie and Stanley Chapman, editors. *Necrologies*. Atlas Press, London, 2007. ISBN 978-1-900565-42-4.

Roger Shattuck. *The Banquet Years*. Faber, London, 1959. ISBN b5917098.

James Anthony Cutshall. *The Figure of the Writer - Alfred Jarry*. Thesis, University of Reading, 1988.

Marieke Dubbelboer. *'UBUSING' CULTURE*. Thesis, Rijksuniversiteit Groningen, 2009.

Zoë Corbyn. An introduction to 'Pataphysics, 2005. URL http://www.guardian.co.uk/culture/2005/dec/09/8/print.

Jean Baudrillard. Pataphysics, 2007. URL http://www.ctheory.net/articles.aspx?id=569.

Asger Jorn. Pataphysics - A Religion In The Making. *Internationale Situationniste*, 6, 1961. URL http://www.infopool.org.uk/6104.html.

Douglas Cruickshank. Why Anti-Matter Matters, nd. URL http://www.ralphmag.org/jarry.html.

Boris Vian. *'Pataphysics? What's That?* Atlas Press, London, 2006. ISBN 1-900565-32-3.

Rene Daumal. *Pataphysical Essays*. Wakefield Press, Cambridge, Massachusetts, 2012. ISBN 978-0-9841155-6-3.

Alistair Brotchie, editor. *A True History of the College of 'Pataphysics - 1*. Atlas Press, London, 1995. ISBN 0-947757-78-3.

Jorge Luis Borges. *Labyrinths - Selected Stories & Other Writings*. New Directions, New York, 1964. ISBN 978-0811200127.

Jorge Luis Borges. *Collected fictions*. Penguin, 1999. ISBN 9780140286809.

Jorge Luis Borges and Margarita Guerrero. *Book of Imaginary Beings*. Viking, 1957.

Jorge Luis Borges and L.S. Dembo. Interview with Borges. *Contemporary Literature*, 11(3):315–323, 2010.

Jorge Luis Borges. *La biblioteca de Babel*. Reclam, 2010. URL http://www.reclam.de/detail/978-3-15-019788-2/Borges__Jorge_Luis/La_biblioteca_de_Babel.

Michel Foucault. The Order of Things - Preface. In *The Order of Things*, chapter Preface, pages xv – xxiv. Editions Gallimard, France, 1966.

Richard E Mayer. Fifty Years of Creativity Research. In Robert J Sternberg, editor, *Handbook of Creativity*, chapter 22, pages 449–460. Cambridge University Press, New York, 1999. ISBN 978-0-521-57604-8.

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

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

Mel Rhodes. An analysis of creativity. *The Phi Delta Kappan*, 42(7):305–310, 1961.

Robert J Sternberg. *Handbook of creativity*. Cambridge University Press, 1999. ISBN 0521576040. URL http://books.google.com/books?id=d1KTEQpQ6vsC&pgis=1.

Arthur Koestler. *The Act of Creation*. Hutchinson & Co, London, 1964. ISBN 0091282705. URL http://www.amazon.co.uk/Creation-Picador-Books-Arthur-Koestler/dp/0330244477/ref=ntt_at_ep_dpt_5.

Henri Poincare. The Value of Science, 2001.

Graham Wallas. *The Art of Thought*. Jonathan Cape; First Edition edition, 1926. URL http://www.amazon.co.uk/art-Thought-Graham-Wallas/dp/B0000654TC/ref=sr_1_1?s=books&ie=UTF8&qid=1320422010&sr=1-1http://books.google.co.uk/books/about/The_art_of_thought.html?id=QO9aAAAAQAAJ.

Margaret Boden. *The Creative Mind: Myths and Mechanisms*. Routledge, London, 2nd edition, 2003. ISBN 0415314534.

Derek Partridge and Jon Rowe. *Computers and Creativity*. Intellect, Oxford, 1994. ISBN 187151651.

George Polya. *How To Solve It*. Princeton University Press, Princeton, New Jersey, 2nd edition, 1957. ISBN 0691023565.

James C. Kaufman and Ronald A. Beghetto. Beyond big and little: The four c model of creativity. *Review of General Psychology*, 13(1):1–12, 2009. ISSN 1939-1552. doi: 10.1037/a0013688. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/a0013688.

Robert J Sternberg. The Nature of Creativity. *Creativity Research Journal*, 18(1):87–98, 2006.

Ronald A. Beghetto and James C. Kaufman. Toward a broader conception of creativity: A case for "mini-c" creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 1(2):73–79, 2007. ISSN 1931-3896. doi: 10.1037/1931-3896.1.2.73. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/1931-3896.1.2.73.

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

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

Marvin Minsky. K-Lines : A Theory of Memory *. *Cognitive Science*, 33(4):117–133, 1980.

Marvin Minsky. *The Society of Mind*. Simon & Schuster, 1988. ISBN 0671657135. URL http://www.amazon.co.uk/Society-Mind-Touchstone-book/dp/0671657135.

David Gelernter. *The Muse in the Machine*. Fourth Estate Limited, London, 1994. ISBN 1-85702-083-9. URL http://www.amazon.com/Muse-Machine-David-Gelernter/dp/0029116023.

Matthew Elton. Artificial Creativity: Enculturing Computers. *Leonardo*, 28(3): 207 – 213, 1995. URL http://www.jstor.org/stable/1576076.

Geraint A Wiggins. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*, 19(7):449–458, November 2006. ISSN 09507051. doi: 10.1016/j.knosys.2006.04.009. URL http://linkinghub.elsevier.com/retrieve/pii/S0950705106000645.

Simon Colton. Creativity versus the perception of creativity in computational systems. In *In Proceedings of the AAAI Spring Symp. on Creative Intelligent Systems*, 2008a.

Graeme Ritchie. Assessing creativity. In *AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science*, pages 3–11. Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science, 2001. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.102.9689&rank=4.

Graeme Ritchie. Some Empirical Criteria for Attributing Creativity to a Computer Program. *Minds and Machines*, 17(1):67–99, June 2007. ISSN 0924-6495. doi: 10.1007/s11023-007-9066-2. URL http://www.springerlink.com/index/10.1007/s11023-007-9066-2.

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

Simon Colton. Computational Creativity. *AISB Quarterly*, pages 6–7, 2008b.

Simon Colton and Geraint A Wiggins. Computational Creativity: The Final Frontier? In *Proceedings of the 20th European Conference on Artificial Intelligence*, pages 21–26, Montpellier, France, 2012. IOS Press. ISBN 978-1-61499-097-0.

Andrew Hugill. Introduction: transdisciplinary learning for digital creative practice. *Digital Creativity*, 24(3):165–167, September 2013. ISSN 1462-6268. doi:

10.1080/14626268.2013.814546. URL http://www.tandfonline.com/doi/abs/10.1080/14626268.2013.814546.

Amnon H. Eden. Three Paradigms of Computer Science. *Minds and Machines*, 17 (2):135–167, July 2007. ISSN 0924-6495. doi: 10.1007/s11023-007-9060-8. URL http://link.springer.com/10.1007/s11023-007-9060-8.

Andrew Hugill and Hongji Yang. The creative turn: new challenges for computing. *International Journal of Creative Computing*, 1(1):4–19, 2013.

Hongji Yang. Editorial. *International Journal of Creative Computing*, 1(1):1–3, 2013.

Johanna Drucker and B Nowviskie. Speculative Computing: Aesthetic Provocations in Humanities Computing. In Susan Schreibman, John Unsworth, and Ray Siemens, editors, *A Companion to Digitial Humanities*, chapter 29. Blackwell Publishing, Oxford, 2007. URL http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml.

Anne Burdick, Johanna Drucker, Peter Lunefeld, Todd Presner, and Jeffrey Schnapp. *Digital_Humanities*. MIT Press, Cambridge, Massachusetts, 2012. ISBN 9780262018470. URL http://mitpress.mit.edu/sites/all/modules/pubdlcnt/pubdlcnt.php?file=/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf&nid=196364.

Sue Thomas, Chris Joseph, Jess Laccetti, Bruce Mason, Simon Mills, Simon Perril, and Kate Pullinger. Transliteracy: Crossing divides. *First Monday*, 12(12), 2007. URL http://journals.uic.edu/ojs/index.php/fm/article/view/2060/1908.

Neil McBride. Robot Ethics: The Boundaries of Machine Ethics, November 2013. ISSN 1095-9203.

Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval: The Concepts and Technology Behind Search*. Addison Wesley, 2011. ISBN 0321416910. URL http://www.mir2ed.org/.

Alistrair Sutcliffe and Mark Ennis. Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10:321–351, 1998.

Erica de Vries. Browsing vs Searching. *OCTO report 93/02*, 1993.

Gary Marchionini. From finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

Baidu. Baidu About, 2012. URL http://ir.baidu.com/phoenix.zhtml?c=188488&p=irol-homeprofile.

Microsoft. Bing Fact Sheet, 2012. URL http://www.microsoft.com/en-us/news/download/presskits/bing/docs/MSBingAll-UpFS.docx.

Sergey Brin and Larry Page. The PageRank Citation Ranking: Bringing Order to the Web. *World Wide Web Internet And Web Information Systems*, pages 1–17, 1998a.

Sergey Brin and Larry Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, April 1998b. ISSN 01697552. doi: 10.1016/S0169-7552(98)00110-X. URL http://linkinghub.elsevier.com/retrieve/pii/S016975529800110X.

Gary Marchionini and Ben Shneiderman. Finding facts vs. browsing knowledge in hypertext systems. *Computer*, 21(1):70–80, January 1988. doi: 10.1109/2.222119.

Christopher Manning, Prabhakar Raghavan, and Hinrich Schuetze. *Introduction to Information Retrieval*. Cambridge UP, online edi edition, 2009. URL http://informationretrieval.org.

Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, September 1999. ISSN 00045411. doi: 10.1145/324133.324140. URL http://portal.acm.org/citation.cfm?doid=324133.324140http://www.cs.cornell.edu/home/kleinber/auth.pdf.

Jon M Kleinberg, Ravi Kumar, Prabhakar Raghavan, and Andrew S Tomkins. The Web as a graph : measurements , models , and methods. *Computer*, 1999.

Krishna Bharat and George Mihaila. Hilltop: A Search Engine based on Expert Documents. In *Proc of the 9th International WWW*, volume 11, 2000. URL ftp://ftp.cs.toronto.edu/pub/reports/csri/405/hilltop.html.

Paul De Bra and Reinier Post. Information retrieval in the World-Wide Web: Making client-based searching feasible. *Computer Networks and ISDN Systems*, 27(2):183–192, November 1994a. ISSN 01697552. doi: 10.1016/0169-7552(94)90132-5. URL http://linkinghub.elsevier.com/retrieve/pii/0169755294901325.

Paul De Bra and Reinier Post. Searching for Arbitrary Information in the WWW : the Fish Search for Mosaic. *Mosaic A Journal For The Interdisciplinary Study Of Literature*, 1994b.

Paul De Bra, Geert-jan Houben, Yoram Kornatzky, and Reinier Post. Information Retrieval in Distributed Hypertexts. *Techniques*, 1994.

M Hersovici, M Jacovi, Y Maarek, D Pelleg, M Shtalhaim, and S Ur. The shark-search algorithm. An application: tailored Web site mapping. *Computer Networks and ISDN Systems*, 30(1-7):317–326, April 1998. ISSN 01697552. doi: 10.1016/S0169-7552(98)00038-5. URL http://linkinghub.elsevier.com/retrieve/pii/S0169755298000385http://www.cs.cmu.edu/~dpelleg/bin/360.html.

Fang-fang Luo, Guo-long Chen, and Wen-zhong Guo. An Improved "Fish-search" Algorithm for Information Retrieval. *2005 International Conference on Natural Language Processing and Knowledge Engineering*, 00:523–528, 2005. doi: 10.1109/NLPKE.2005.1598793. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1598793.

Carmelo Bastos Filho, Fernando de Lima Neto, Anthony Lins, Antonio Nascimento, and Marilia Lima. A novel search algorithm based on fish school behavior. *IEEE International Conference on Systems, Man and Cybernetics*, pages 2646–2651, October 2008. ISSN 1062-922X. doi: 10.1109/ICSMC.2008.4811695. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4811695http://www.fbln.pro.br/fss/.

Z.Z. Nick and P. Themis. Web Search Using a Genetic Algorithm. *IEEE Internet Computing*, 5(2):18–26, 2001. ISSN 10897801. doi: 10.1109/4236.914644. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=914644.

Eugene Agichtein, Eric Brill, and Susan Dumais. Improving web search ranking by incorporating user behavior information. In *ACM SIGIR conference on Research and development in information retrieval*, page 19, New York, New York, USA, 2006. ACM Press. ISBN 1595933697. doi: 10.1145/1148170.1148177. URL http://portal.acm.org/citation.cfm?doid=1148170.1148177.

Shenghua Bao, Xiaoyuan Wu, Ben Fei, Guirong Xue, Zhong Su, and Yong Yu. Optimizing Web Search Using Social Annotations. *Distribution*, pages 501–510, 2007.

Hector Garcia-molina, Jan Pedersen, and Zoltan Gyongyi. Combating Web Spam with TrustRank. In *In VLDB*, pages 576–587. Morgan Kaufmann, 2004.

E.J. Glover, G.W. Flake, Steve Lawrence, W.P. Birmingham, A. Kruger, C.L. Giles, and D.M. Pennock. Improving category specific Web search by learning query modifications. *Proceedings 2001 Symposium on Applications and the Internet*, pages 23–32, 2001. doi: 10.1109/SAINT.2001.905165. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=905165.

Taher H Haveliwala. Topic-Sensitive PageRank : A Context-Sensitive Ranking

Algorithm for Web Search. *Knowledge Creation Diffusion Utilization*, 15(4):784–796, 2003.

Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Information retrieval in folksonomies: Search and ranking. In *The Semantic Web: Research and Applications, volume 4011 of LNAI*, pages 411–426. Springer, 2006.

Glen Jeh and Jennifer Widom. Simrank: A measure of structural-context similarity. In *In KDD*, pages 538–543, 2002.

Bo Shu and Subhash Kak. A neural network-based intelligent metasearch engine. *Information Sciences*, 120, 1999.

D.H. Widyantoro and J. Yen. A fuzzy ontology-based abstract search engine and its user studies. *10th IEEE International Conference on Fuzzy Systems. (Cat. No.01CH37297)*, 2:1291–1294, 2001. doi: 10.1109/FUZZ.2001.1008895. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1008895.

Zhi-Qiang Du, Jing Hu, Hong-Xia Yi, and Jin-Zhu Hu. The Research of the Semantic Search Engine Based on the Ontology. *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, pages 5398–5401, September 2007. doi: 10.1109/WICOM.2007.1323. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4341098.

Li Ding, Tim Finin, Anupam Joshi, Rong Pan, R. Scott, Cost Yun Peng, Pavan Reddivari, Vishal Doshi, and Joel Sachs. Swoogle: A semantic web search and metadata engine. In *In Proceedings of the 13th ACM Conference on Information and Knowledge Management. ACM*, 2004.

Jaap Kamps, Rianne Kaptein, and Marijn Koolen. Using anchor text , spam filtering and wikipedia for web search and entity ranking. Technical report, ?, 2010.

Mohammad Mustafa Taye. *Ontology Alignment Mechanisms for Improving Web-based Searching*. PhD thesis, De Montort University, 2009.

Sadaaki Miyamoto. *Fuzzy Sets in Information Retrieval and Cluster Analysis (Theory and Decision Library D:)*. Springer, 2010. ISBN 9048140676. URL http://www.amazon.co.uk/Information-Retrieval-Cluster-Analysis-Decision/dp/9048140676.

Sadaaki Miyamoto. Information Retrieval based on Fuzzy Associations, 1988.

P Srinivasan. Vocabulary mining for information retrieval: rough sets and fuzzy sets. *Information Processing & Management*, 37(1):15–38, January 2001. ISSN

03064573. doi: 10.1016/S0306-4573(00)00014-5. URL http://linkinghub.elsevier.com/retrieve/pii/S0306457300000145.

Sadaaki Miyamoto and K Nakayama. Fuzzy Information Retrieval Based on a Fuzzy Pseudothesaurus. *IEEE Transactions on Systems, Man and Cybernetics*, 16(2):278–282, 1986.

Scott Deerwester, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by Latent Semantic Analysis. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*, 41(6):391–407, 1990.

Craig Macdonald. The Voting Model for People Search. *Philosophy*, 2009.

Hinrich Schuetze. Automatic Word Sense Discrimination. *Computational Linguistics*, 1998.

Hinrich Schuetze and Jan Pedersen. Information retrieval based on word senses, 1995.

Jeffrey Dean, Luiz Andre Barroso, and Urs Hoelzle. Web Search for a Planet: The Google Cluster Architecture. *Ieee Micro*, pages 22–28, 2003.

Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, Sebasopol, CA, 1st edition, 2009. ISBN 978-0-596-51649-9.

Daniel Jurafsky and James H Martin. *Speech and Language Processing*. Pearson Education, London, 2nd edition, 2009. ISBN 978-0-13-504196-3.

Mitchell P Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a Large Annotated Corpus of English : The Penn Treebank. *Computational Linguistics*, 19(2), 1993.

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

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

Alison Pease, Simon Colton, Ramin Ramezani, John Charnley, and Kate Reed. A Discussion on Serendipity in Creative Systems. In *Proceedings of the 4th International Conference on Computational Creativity*, volume 1000, pages 64–71, Sydney, Australia, 2013. University of Sydney. ISBN 9781742103174. URL http://www.computationalcreativity.net/iccc2013/.

Davide Piffer. Can creativity be measured ? An attempt to clarify the notion of creativity and general directions for future research. *Thinking Skills and Creativity*, 7(3):258–264, 2012. ISSN 1871-1871. doi: 10.1016/j.tsc.2012.04.009. URL http://dx.doi.org/10.1016/j.tsc.2012.04.009.

Warren Motte. *Oulipo, A primer of potential literature*. Dalkey Archive Press, London, 2007. ISBN 1-56478-187-9.

Kenneth M Heilman, Stephen E Nadeau, and David O Beversdorf. Creative innovation: possible brain mechanisms. *Neurocase*, 9(5):369–79, October 2003. ISSN 1355-4794. doi: 10.1076/neur.9.5.369.16553. URL http://www.ncbi.nlm.nih.gov/pubmed/14972752.

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

Bernd Carsten Stahl, Marina Jirotka, and Grace Eden. Responsible Research and Innovation in Information and Communication Technology: Identifying and Engaging with the Ethical Implications of ICTs. In Richard Owen, editor, *Responsible Innovation*, chapter 11, pages 199–218. John Wiley and Sons, 2013.

Glossary

bisociation
Two self-consistent but habitually incompatible frames of reference intersecting to give rise to a new creative idea.. 39

holonym
The relationship between a term denoting the whole and a term denoting a part of, or a member of, the whole. That is, 'X' is a holonym of 'Y' if Ys are parts of Xs, or 'X' is a holonym of 'Y' if Ys are members of Xs. For example, 'tree' is a holonym of 'bark', of 'trunk' and of 'limb.' Holonymy is the opposite of meronymy.. 93

hypernym
A hyponym shares a type-of relationship with its hypernym. For example, pigeon, crow, eagle and seagull are all hyponyms of bird (their hypernym); which, in turn, is a hyponym of animal.. 93