
Mane, tekel, fares

Perspektywy metod
obliczeniowych w historii
(pojęć), czyli o
barbarzyńcach z ekierkami

Plan

1. Wprowadzenie
2. Historia mierzenia (w) historii
3. Co to jest historia obliczeniowa?
 - a. Opis i przykłady
 - b. Przykład 1 - Paul Thagard i przełomy pojęciowe w historii nauki*
 - c. Przykład 2 - Arianna Betti (et al.) i "pojęcia w ruchu"*
 - d. Przykład 3 - lokalny prospekt grantowy*
4. Co z tego wynika? Wnioski, podsumowania, otwarte problemy

1. Dlaczego i po co kwantyfikacja?

- od historii kwantyfikacji do kwantyfikacji (w) historii
- ukierunkowanie: historia pojęć w historii nauki i historii filozofii
 - specyfika HPS? (history and philosophy of science) - systematyczność, konwencjonalność, dbałość o ścisłość, dyscyplinarność, bliskość historii i filozofii nauki
 - odrębność rozmywana: różnica stopnia niż rodzaju
 - przekładalność? -> np. pojęcia społeczno polityczne
- historia pojęć jako historia systemów poznawczych (epistemologia historyczna);
 - kontekst kognitywistyczny (akwizycja, stosowanie i "przemeblowywanie" pojęć)
- Cele: odczarowanie tematu
 - spóźnienie historii - nawet jeśli nie stosujemy, powinniśmy rozumieć
 - nie budujemy maszynek
 - nie jest to jednolita, normatywna metodologia; żaden scjentyzm-pozytywizm!
 - to tylko narzędzia - różna skala stosowalności i sensowności stosowania - ale jak szeroka?
 - w najgorszym wypadku - wzmocnienie sceptyków

2. Historia kwantyfikacji w historiografii

- Archeologia: Buckle i dyskusje statystyczne XIX w.
- Kliometria (od lat 60. XX w.)
 - przełom ekonometryczny
 - New Economic History - zob. Boldizzoni 2011
- Kliodynamika (XXI w.)
 - makroskala, procesy dynamiczne
 - poszukiwanie cykli - nędza historycyzmu?
 - Maddison Project
 - Clio Infra
 - Seshat Global Database (Peter Turchin)
- Historia obliczeniowa (pojęć)
 - modelowanie sieci
 - Moretti (2003)
 - Grasshoff
 - Armitage; Guldi

3.1.1. Obliczeniowa historia

- najogólniej: aplikacja istniejących i znakomicie działających narzędzi z technologii przetwarzania języka naturalnego (NLP) do problemów historycznych
- przeznaczone do zadań wymagających pracy z dużym korpusem
- **zadaniem jest *pomoc* historykowi - nie zastąpienie go**
- **“measurement is evidence but not sole evidence”**
- główne przykazanie *data science*: poszukuj wzorca
- źródła - jako dane (ale nie tylko!)
- *close reading a distant reading*
- patrzymy na korpusy danych, tłumaczymy na język liczb (mierzenie albo kwantyfikacja)
- cyfrowa obliczeniowa
 - ditigitalizacja -> digitalizacja + analiza



3.1.2. Obl. hist. systemów pojęciowych

- idea: pojęcia działają w systemach, nabierają i zmieniają znaczenie we wzajemnych relacjach - kontekstach (hipoteza dystrybucyjna Firtha)
- można mapować te schematy (np. jako odległości w przestrzeni)
- narzędziami można się wspierać albo operacjonalizować problemy - nigdy jednak nie jest to eksperyment
- procedura: budowanie bazy danych -> ew. wygładzanie danych -> (budowanie ontologii [modelowanie]) -> wizualizacja, analiza albo testy
- procedura przypomina nieco koło hermeneutyczne: sprzężenie zwrotne (założenia+reguły<->wnioski => akceptacja albo odrzucenie; ew. udoskonalanie zał./reg. w przypadku wykrycia błędu -
 - w historii obliczeniowej to koło to współpraca eksperta z narzędziami

3.1.3. Do czego się nadaje?

- Jest jasne, że nie do wszystkiego - różna podatność na kwantyfikację
- Specyfika historii - dziurawość korpusów, często niepewność danych, "zamkniętość" materii jaką jest przeszłość
- Uzależnienie nie tylko od dziedziny, tematu, ale i stawianego pytania
 - ->Podejście pragmatyczne: pluralizm w pluralistycznej dziedzinie
 - "Multiple your ways of knowing" (D.H. Fisher 1970)
- Szczególnie dobrze - mapy zależności i przepływów - kontakty naukowe, schematy idei, stosunki gospodarcze, stosunki przestrzenne (migracje)
 - transmisja wiedzy - jak szybko podróżują fakty (Howlett & Morgan 2011)

cd. - płynny problem

- Wciąż młoda dziedzina - wiele projektów jest wekslami albo tylko "proofs-of-concept"
- Ta granica się przesuwa - rozwój metod: OCR, moc obliczeniowa, wielkość korpusów, narzędzia analityczne (przetwarzanie maszynowe, półautomatyczne)
- Opór mentalny słabnie - obserwacja sytuacji w np. naukach społecznych
- Wsparcie techniczne niezbędne - wszechstronność historyka (np. język programowania lub znajomość narzędzi) i/lub interdyscyplinarny zespół

Kluczowe pytanie: czy potrafią coś więcej niż tradycyjne metody? czy nie dowieemy się najwyżej tego, co już wiemy?

3.1.4. Przykłady sukcesów (Mullen 2018)

“careful empirical work to overturn conventional wisdom”

- liczne w historii nauki - John Snow, Galton
- Arbuthnot i Zabell
- Fogel - Railroads and the economic growth (1964)
- Mobilność społeczna w Bostonie (Thernstrom 1964)
- Fogel i Engerman - Time on the Cross (1974)
- Mentalna geografia w Houston (Blevins -
<http://web.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=93>)
- Viral Texts (Cordell 2017 - <http://viraltxts.org/>)
- American Bible (Mullen 2017 - <http://americaspublicbible.org/>)
- Old Bailey Proceedings (Hitchcock & Turkel 2016 -
<https://www.oldbaileyonline.org/>)

3.1.5 Metody

-> różny stopień “autonomii” i złożoności analizy

-> metody wstępujące i zstępujące (obie obciążone założeniami)

- analiza frekwencyjna: n-gramowa (np. Google Ngram) - Michel et al. 2011 - *culturomics*
 - *Guldi 2012*
- semantyka dystrybucyjna
 - latent semantic analysis (LSA)
 - reprezentacja wektorowa
 - miary podobieństwa *tekstów*
- przeszukiwalne bazy odpowiadające na konkretne zapytania i segregujące dane według zadanych zmiennych

Metody - cd.

- klastrowanie
- “wychwytywanie” danej kategorii z tekstu (*data mining* + *kategoryzacja*)
- analiza metadanych
- anotowanie kontekstowo relewantnych miejsc (czwórki uporządkowane - Vagon)
- wizualizacje
 - mapy
 - wykresy
 - grafy
 - diagramy
- interaktywność - aktywny użytkownik i ew. odbiorca

3.2 Paul Thagard

Zainteresowanie **przełomami poznawczymi**
tradycja Kuhnowska filozofii nauki

Przekonanie, że problem jest kognitywistyczny



- Przekonanie, że intuicje są niesatysfakcjonujące metodologicznie - chęć formalizacji -> odwołanie do metod ontologii formalnej

Zastosowanie programu komputerowego (ECHO) - modelowanie: wyznaczanie części danych systemów pojęciowych (wierzchołki) i relacji między nimi (krawędzie);

Teoria naukowa - koherentna eksplanacyjnie - różnice tej koherencji można porównywać i tłumaczyć w ten sposób, dlaczego jedna jest lepsza od drugiej

3.2.1 ECHO

Rozpisanie struktury
logicznej teorii
Lavoisiera i teorii
flogistonowej jako
rywalizujących
wyjaśnień.

Table I. *Input propositions for Lavoisier (1862) example*

Evidence

- | | |
|------------------|--|
| (proposition 'E1 | "In combustion, heat and light are given off.") |
| (proposition 'E2 | "Inflammability is transmittable from one body to another.") |
| (proposition 'E3 | "Combustion only occurs in the presence of pure air.") |
| (proposition 'E4 | "Increase in weight of a burned body is exactly equal to weight of air absorbed.") |
| (proposition 'E5 | "Metals undergo calcination.") |
| (proposition 'E6 | "In calcination, bodies increase weight.") |
| (proposition 'E7 | "In calcination, volume of air diminishes.") |
| (proposition 'E8 | "In reduction, effervescence appears.") |

Oxygen hypotheses

- | | |
|-------------------|---|
| (proposition 'OH1 | "Pure air contains oxygen principle.") |
| (proposition 'OH2 | "Pure air contains matter of fire and heat.") |
| (proposition 'OH3 | "In combustion, oxygen from the air combines with the burning body.") |
| (proposition 'OH4 | "Oxygen has weight.") |
| (proposition 'OH5 | "In calcination, metals add oxygen to become calxes.") |
| (proposition 'OH6 | "In reduction, oxygen is given off.") |

Phlogiston hypotheses

- | | |
|-------------------|--|
| (proposition 'PH1 | "Combustible bodies contain phlogiston.") |
| (proposition 'PH2 | "Combustible bodies contain matter of heat.") |
| (proposition 'PH3 | "In combustion, phlogiston is given off.") |
| (proposition 'PH4 | "Phlogiston can pass from one body to another.") |
| (proposition 'PH5 | "Metals contain phlogiston.") |
| (proposition 'PH6 | "In calcination, phlogiston is given off.") |

Table 2. *Input explanations and contradictions in Lavoisier (1862) example*

Oxygen explanations

- (explain '(OH1 OH2 OH3) 'E1)
- (explain '(OH1 OH3) 'E3)
- (explain '(OH1 OH3 OH4) 'E4)
- (explain '(OH1 OH5) 'E5)
- (explain '(OH1 OH4 OH5) 'E6)
- (explain '(OH1 OH5) 'E7)
- (explain '(OH1 OH6) 'E8)

Phlogiston explanations

- (explain '(PH1 PH2 PH3) 'E1)
- (explain '(PH1 PH3 PH4) 'E2)
- (explain '(PH5 PH6) 'E5)

Contradictions

- (contradict 'PH3 'OH3)
- (contradict 'PH6 'OH5)

Data

- (data '(E1 E2 E3 E4 E5 E6 E7 E8))

oxygen and phlogiston theories, respectively. These propositions do not capture Lavoisier's argument completely but do recapitulate its major points. (In a slightly more complicated simulation not presented here, I have encoded the attempt by the phlogiston theory to explain the increase in weight in combustion and calcination by

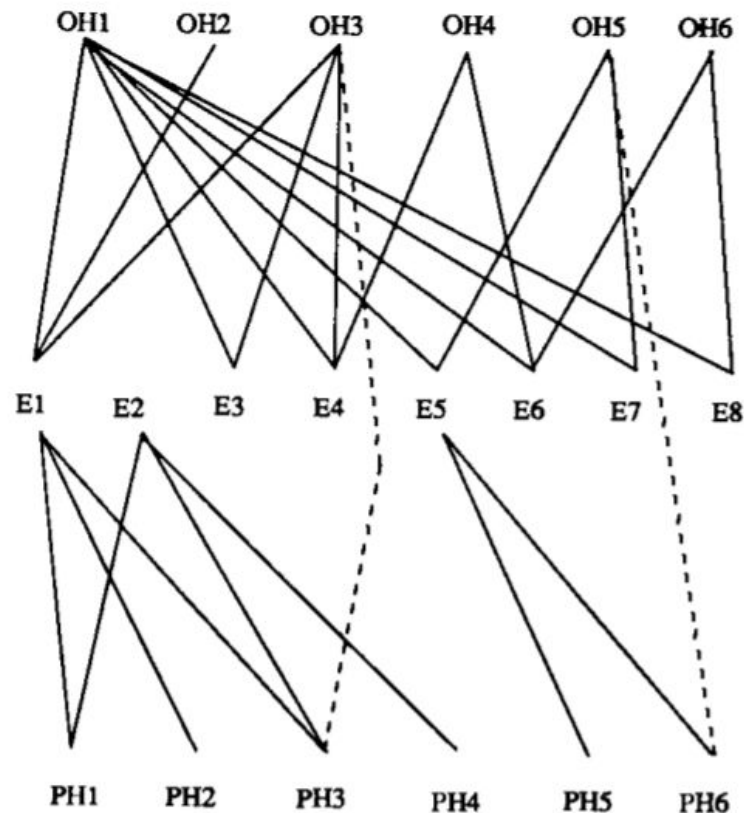


Figure 9. Network representing Lavoisier's (1862) argument. E1-E8 are evidence units. OH1-OH6 are units representing hypotheses of the oxygen theory; PH1-PH6 represent the phlogiston hypotheses. Solid lines are excitatory links; dotted lines are inhibitory.

3.2.2 Schematy jako repr. str. log. teorii naukowej

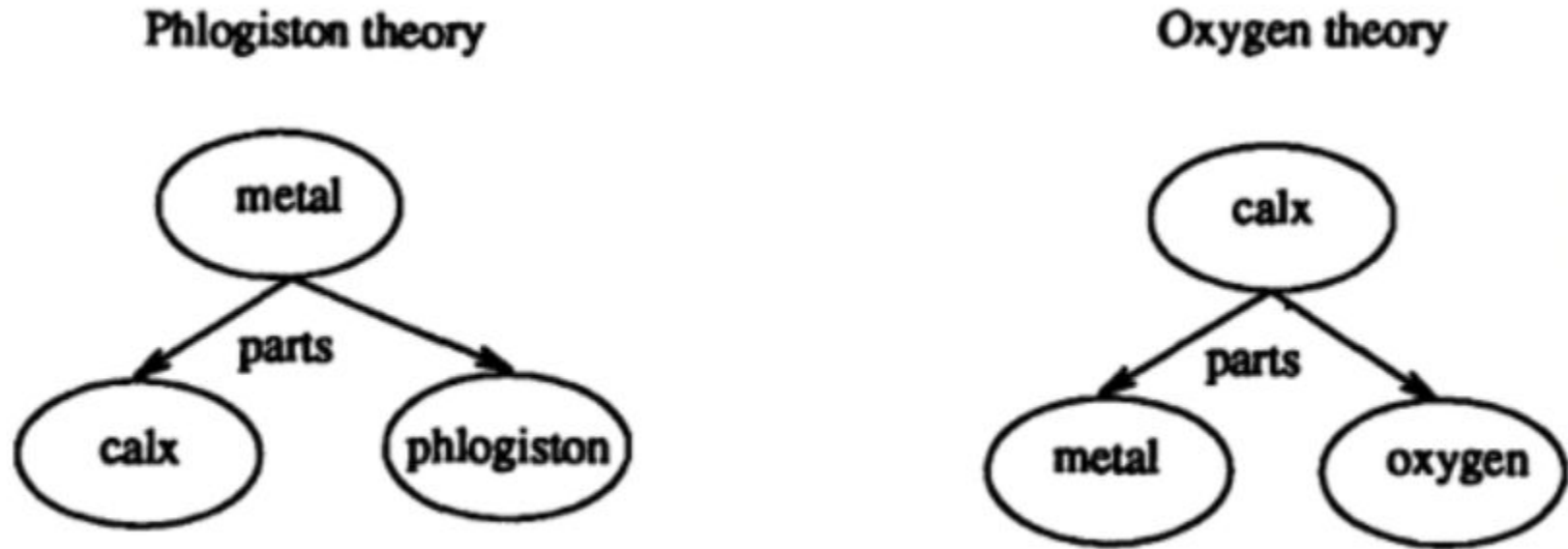


Figure 3.7. Part-relations in the phlogiston and oxygen theories.

3.3 Arianna Betti

Betti, v.d. Berg: "Toward A Computational History of Ideas" (2014): "Sound method + organised data"

metody "wstępujące" (top-down) (i "zstępujące" (bottom-up)

Śledzenie "concept-units", ale z rozróżnieniem na stały "rdzeń" i zmienny "margines" - przykład: "Classical Axiomatic Science"

Stworzenie narzędzia SalVe do pomocy historykom filozofii

SalVe

- Użytkownik wprowadza do programu dane (teksty) podzielone (arbitralnie) na „dokumenty”. Mogą nimi być zdania, paragrafy, rozdziały, całe książki.
- Dokumenty rozbijane są na *terminy* (pojedyncze jednostki semantyczne), które są *normalizowane* a następnie informacja o nich przechowywana jest w pliku indeksu.
- SalVe wykorzystuje ideę reprezentacji wektorowej dokumentów (*Vector Space Model*). Oznacza to, że każdy dokument przechowywany jest w postaci wektora częstości występowania słów kluczowych.

Salve c.d.

wyszukiwanie dokumentów zawierających podane terminy w bazie

porównywanie dokumentów ze względu na ich podobieństwo

wyszukiwanie najbardziej podobnych dokumentów do określonego dokumentu

podgląd najbardziej "charakterystycznych" dla danego dokumentu terminów

eksport danych dotyczących częstości występowania słów w bazie

eksport danych dotyczących podobieństwa między dokumentami w formacie umożliwiającym wizualizację

3.3.1

Porównywanie dokumentów

Dokument A: Świat jest wszystkim [=wszystko], co jest faktem [=fakt].

Dokument B: Świat jest ogółem [=ogół] faktów [=fakt], nie rzeczy [=rzecz].

Mamy więc pięć terminów: "Świat", "wszystko", "fakt", "ogół", "rzecz".

Możemy więc nasze dwa dokumenty reprezentować jako macierz:

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (1)$$

A następnie obliczyć podobieństwo między nimi:

$$\text{sim}(a, b) = \frac{V(a) \cdot V(b)}{\|V(a)\| \cdot \|V(b)\|} \quad (2)$$

Dodatkowo częstość występowania terminów w dokumencie skorygowana jest zgodnie z zasadą *Term Frequency Inverse Document Frequency*, to znaczy dla każdego terminu w dokumencie jego częstość wystąpień podzielona jest przez liczbę dokumentów, w których dany termin występuje. Pozwala to zminimalizować wagę "niecharakterystycznych" terminów.

3.3.2 Stawianie hipotez

Teza interpretacyjna: Bolzano twierdzi, że każda *analityczna* prawda jest ugruntowana w jakiejś prawdzie *syntetycznej* (de Jong 2001, s. 346)

1. Co Bolzano pisał o *analityczności* oraz relacji *ugruntowania*?
2. Jak te dwa pojęcia wpisują się w ogólny obraz *Wissenschaftslehre*?
3. Czy są jakieś świadectwa na to, że Bolzano łączył dystynkcje *analityczne-syntetyczne* z pojęciem ugruntowania.

3.3.3 Stosowanie narzędzia SalVe

Jak przeformułować te pytania do postaci, w której mogą być przedmiotem analizy w programie SalVe?

1. W których paragrafach słowo *analytisch* pojawiało się najczęściej? (zaawansowane wyszukiwanie)
2. Które paragrafy są najbardziej podobne do tych, w których *analytisch* pojawia się najczęściej? (obliczanie podobieństwa)
3. Które słowa współwystępują z *analytisch* najczęściej? (własności terminu, kolokacje)
4. W jaki sposób *lehre* oraz *lehrsatz* i *abfolge* rozkładają się w całości dzieła Bolzano?

3.4 Projekt Ajdukiewicz

- budowa narzędzia - Sokratyzer
- otwarty dla użytkowników
 - tworzenie własnych korpusów
 - eksport danych - tworzenie wizualizacji
- aplikacja idei zespołu Betti
- analiza idei filozoficznych SL-W; case study: “znaczenie” u Ajdukiewicza
- czy pojęcie zmieniało swoje znaczenie? lub konotacje? jaka jest genealogia tego pojęcia na tym wycinku intelektualnej mapy filozofii analitycznej?
- operacjonalizacja i weryfikacja niektórych tez z literatury przedmiotu

3.4.1 Sokratyzer

Sokratyzer

PlikWidokUstawienia

Wyszukiwanie

Podobieństwo

More Like This

Słowa kluczowe

Wybierz do porównania:

54

55

56

57

58

59

60

Nr paragrafu	Podobieństwo
59	0.017
38	0.002
54	0.041
58	1.0

XLIII. There are also idols formed by the reciprocal intercourse and society of man with man, which we call idols of the market, from the commerce and association of men with each other; for men converse by means of language, but words are formed at the will

XLIV. Lastly, there are idols which have crept into men's minds from the various dogmas of peculiar systems of philosophy, and also from the perverted rules of demonstration, and these we denominate idols of the theatre: for we regard all the systems of

Porównaj z:

54

55

56

57

58

59

60

Porównaj

4.1 Problemy i ograniczenia

- zakres stosowalności w semantyce - wciąż niejasny; raczej tylko asysta
- ograniczenia konkretnych metod: np. LSA i polisemia
- niesamodzielne
- ręczne anotowanie
- problem metadanych (n-gramy - Pechenick et al. 2015)
- tak dobre, jak dane, zestawy treningowe i/lub początkowe klasyfikacje
- **modele nigdy nie są bez założeń**
- wciąż arbitralne (choć potencjalnie mniej arbitralne)
- słabo stosowalne dla klasy problemów - np. wymagających *close reading* lub pracy na małej ilości materiału
- wymagają dodatkowych kompetencji, czasem też dużych nakładów
- problemy techniczne - brak baz, słaba jakość skanów (gotyk myli bazy - "suck") - potrzeba specjalistycznych narzędzi

4.2 Zalety

- Ułatwia
 - zmniejszają obciążenie poznawcze (delegowanie zadania) - dowody kognitywistyczne
 - zmniejsza ryzyko błędu w żmudnej pracy
 - rośnie szansa na zainteresowanie (badaczy i słuchaczy) "technicznym" tematem
- Przyspiesza
 - wzrost skali
 - nawet jeśli człowiek może wykonać
- Zmienia perspektywę (dodaje nową)
 - wzrost skali powoduje możliwość odkrycia struktury
 - pluralizm metod i wyjaśnień jest pożądany - specyfika metodologii historii
- Zwiększa bazę potencjalnych źródeł (casus houstoński) lub kontekstów (Huijnen et al. 2013)
- Uelastycznia
 - mała zmiana nie wprowadza konieczności przepisywania wszystkiego od nowa

4.2 cd. - więcej zalet!

- Wspomaga obserwacje, wnioskowanie i odkrycia
- Umożliwia obserwacje, wnioskowanie i odkrycia
- Uwidacznia strukturę wnioskowania historycznego (zmusza do formalizacji)
 - przesunięcie w prawo na osi "obiegowa opinia/opinia ekspercka <-> intersubiektywna wiedza";
 - transparentność!
 - replikowalność?
- Pozwala sprawdzać (w ogr. zakresie) - wizualizacja łowi błędy (Arbuthnot - Zabell)
 - i siebie
 - i innych

4.3 Co z tego wynika dla metodologii?

- Czy metody obliczeniowe mogą ograniczać intuicję danego badacza
 - zmuszenie do transparentności i intersubiektywności
 - uchylanie czarnej skrzynki, jaką jest praca historyka i procesy w jego głowie
 - zaostrowanie kryteriów - ramy porównawcze dane z góry
 - ograniczenie możliwości uciekania od przyznania racji i dodawania epicykli
- Co jest wystarczającym dowodem?
 - jak się dowodzi w historii? czy zawsze się tylko przekonuje?
 - z liczbami trudno się dyskutuje
 - czy możliwe jest bardziej zobiektywizowane mierzenie "mocy" dowodu?
 - liczenie ilości dowodów - intuicja vs. frekwencja
 - dowody "tradycyjne" z wykorzystaniem statystyk - np. trendy społeczne
- "odwracanie" - (Kuhna: struktura/historia; kognitywistyki: nauka/filozofia)

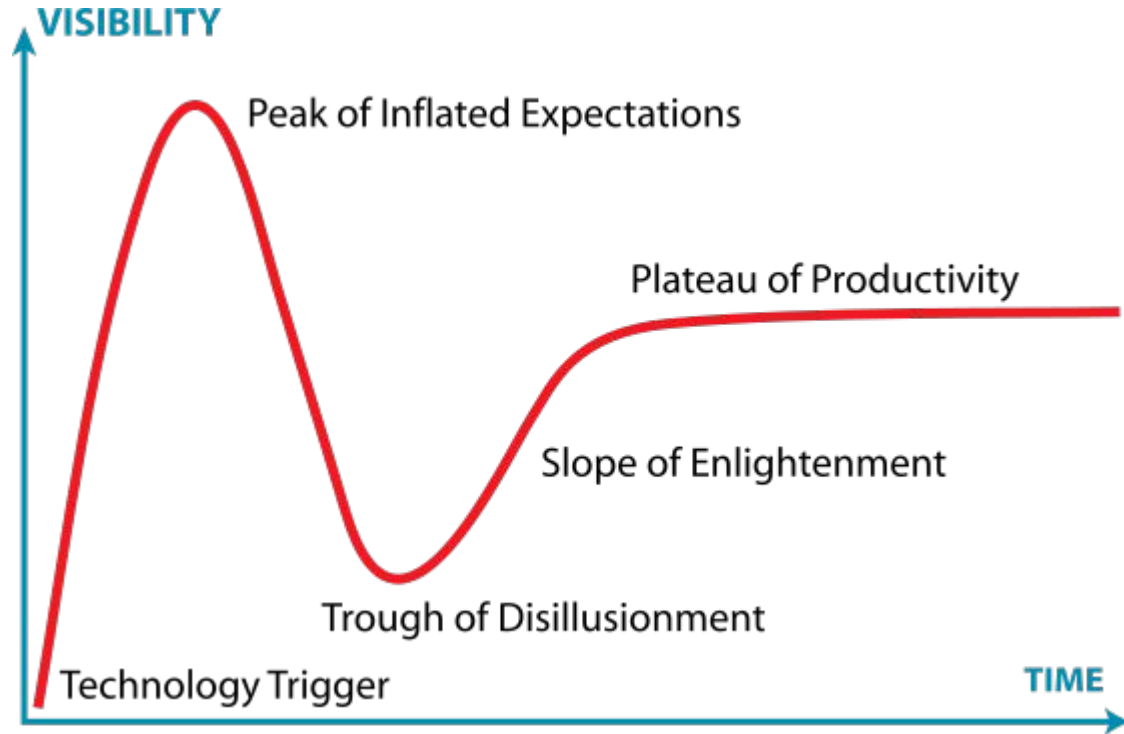
Pytania metodologiczne - cd.

- Jak rozróżnić ograniczenia logiczne ("fatal flaws") od technicznych/usuwalnych
 - czy odkrywamy to, co już wiemy?
 - nie oceniać po błędnych i nietrafionych zastosowaniach
 - przeciwnicy, jeśli chcą odeprzeć barbarzyńców raz na zawsze, muszą wskazać przesłanki logiczne
 - zwolennicy, jeśli chcą przekonać niedowiarków, muszą pokazać, że metody oblicz działają i to lepiej niż istniejące metody - pozwalają zauważać nowe rzeczy albo lepiej udowadniać już znane (lub inaczej odkrywalne)
- Dynamika badacz-narzędzia (system - "poznanie rozszerzone")
 - delegowanie części zadań (np. wybór źródeł albo wstępna analiza)
 - niejasność relacji rośnie wraz z algorytmizacją
 - umiejętność łączenia intuicji z regułami
 - konieczność stałego monitorowania wyników
 - decydowanie, czy dany wynik każe zmienić strukturę wnioskowania, czy odrzucić wynik (por. koło hermeneutyczne)

4.4 Jak podchodzić do historii obliczeniowej?

- Nie zamykać w hasłach - to nie jest “scjentyzm”, “pozytywizm”, “materializm historyczny”, ani “strukturalizm”
- To nie jest jednolita metodologia - a przybornik metod o zmiennej stosowalności - w zależności od problemu i decyzji badacza
- Nie podpalać się - mimo pozorów i języka wojna nie jest kulturowa, frakcyjna, ani religijna (doświadczenie lat 60.)
- Młoda “dyscyplina” - trudno powiedzieć, na co ją stać. Unikać odrzucenia a priori
- Opór nie może wynikać z przyzwyczajenia albo podniebienia
- Podstawą jest właściwe zrozumienie istoty metod obliczeniowych, nawet jeśli dalej chce się je krytykować
- Prawo Amary, krzywa Gartner i drugie Prawo Clarke’a

Gartner Hype Cycle



Niektóre linki

- Project Maddison:
<https://www.rug.nl/ggdc/historicaldevelopment/maddison/>
- Grupa Betti: <https://conceptsinnmotion.org/>
- Indiana Ontology Project <https://inpho.cogs.indiana.edu>
- Seshat: <http://seshatdatabank.info/>
- Oksfordczycy: <http://blogs.bodleian.ox.ac.uk/digital/2018/02/02/dahs2/>
- Nowy żurnal branżowy: <https://jhnr.uni.lu/index.php/jhnr>
- Piękny przykład interaktywnej wizualizacji - środowisko Bauhausu:
<https://www.youtube.com/watch?v=3DB9qU9tK3Y>

Dziękuję za uwagę!

P.S. Bibliografia dla chętnych dostępna - proszę o kontakt.