

# Analiza zbioru Open Academic Graph (OAG)

Kacper Skelnik, Wojciech Tyczyński

2020



# Spis treści

1	Informacje wstępne	2
2	Analiza danych ze zbioru AMiner	2
3	Analiza danych ze zbioru MAG	5
4	Wnioski	9

## 1 Informacje wstępne

Zbiór Open Academic Graph (OAG) został stworzony przez połączenie dwóch dużych zbiorów z publikacjami naukowymi:

- [Microsoft Academic Graph \(MAG\)](#)
- [AMiner](#)

Celem analizy było sprawdzenie, czy korzystając z wymienionych wyżej zbiorów danych możliwe jest wyznaczenie trendów w nauce na przestrzeni lat. W obrębie tego sprawozdania zostanie omówiona analiza tylko części z dostępnych danych. Poniżej znajduje się opis poszczególnych zbiorów oraz metoda ich analizy. Różnice wynikają z innej budowy oraz zawartości każdego z nich. Na końcu sprawozdania znajdują się tabele zbiorcze zawierające esencję wykonanych badań. Wszystkie operacje związane z opracowaniem danych zostały wykonane w języku Python. W czasie pracy używany był serwer Pracowni Fizyki w Ekonomii i Naukach Społecznych Wydziału Fizyki Politechniki Warszawskiej, za którego udostępnienie serdecznie dziękujemy.

## 2 Analiza danych ze zbioru AMiner

AMiner jest zbiorem stworzonym na Tsinghua University jako serwis do wyszukiwania i analizowania akademickiej sieci publikacji naukowych. W skład OAG wchodzi 172,209,563 publikacji ze zbioru AMiner, jest to stan tego zbioru w styczniu 2019 roku. Warto zaznaczyć, że w momencie prowadzenia projektu bezpośrednio w zbiorze AMiner dostępne było już 271,795,672 publikacji i liczba ta rośnie sukcesywnie co kilka sekund.

Największą trudnością w analizie AMiner jest rozmiar plików, z których składa się zbiór jak również zróżnicowanie zapisanych danych o poszczególnych publikacjach. O wielu plikach są wyłącznie podstawowe informacje, takie jak tytuł i autorzy, podczas gdy wiele plików zawiera obszerne abstrakty i wiele dodatkowych informacji.

### Wczytanie danych

Ze względu na duży rozmiar plików, na samym początku opracowany został przy wykorzystaniu biblioteki `os` w pythonie skrypt, którego zadaniem było rozpakowanie z archiwów `zip` plików tekstowych, a następnie podzielenie ich za pomocą linuxowej komendy `sed -n` na mniejsze pliki zawierające po około milion wierszy. Każdy z wierszy jest zapisany w formacie json dlatego dane trzeba wczytywać wiersz po wierszu przy pomocy funkcji `json.loads()` z biblioteki `json`. Wczytywanie wiersz po wierszu jest również w tym przypadku jedną z najwydajniejszych metod wczytywania danych.

### Przygotowanie danych

W projekcie wykorzystane są informacje dotyczące roku wydania oraz słów kluczowych (keywordów), opisujących publikację. Kolumny zawierające inne informacje niż te, które są wykorzystywane

w projekcie zostały usunięte. Rok wydania publikacji zapisany jest w formacie liczby całkowitej, natomiast kolumna zawierająca informacje o słowach kluczowych złożona jest z różnego rozmiaru list, zawierających keywordy. Następnie zostały usunięte wiersze, w których brakowało danych dotyczących roku wydania lub keywordów. Kolumna zawierająca listy keywordów została zamieniona na ramkę danych (ang. *data frame*), w której każda kolumna zawierała pojedynczy keyword, a różnice w rozmiarach list zostały rozwiązane poprzez uzupełnienie list do największego rozmiaru wartościami NA (ang. *not available*, oznaczenie brakującej wartości). Na koniec data frame z keywordami został połączony z kolumną zawierającą lata wydania publikacji oraz przekształcony do postaci 2 kolumnowego data frame gdzie w jednej kolumnie znajdował się rok wydania publikacji, a w drugiej keyword. Data frame został przekształcony do takiej postaci za pomocą funkcji `melt` z biblioteki `pandas`. Przekształcenie danych wejściowych do pożądanej postaci wymagały bardzo dużych zasobów mocy, pamięci i czasu ze względu na rozmiary danych i złożoność przekształceń. Cały proces został zautomatyzowany poprzez przygotowanie odpowiednich funkcji, które były ręcznie testowane na wcześniej przygotowanych zbiorach próbnych które zawierały po około 1000 wierszy.

## Analiza danch

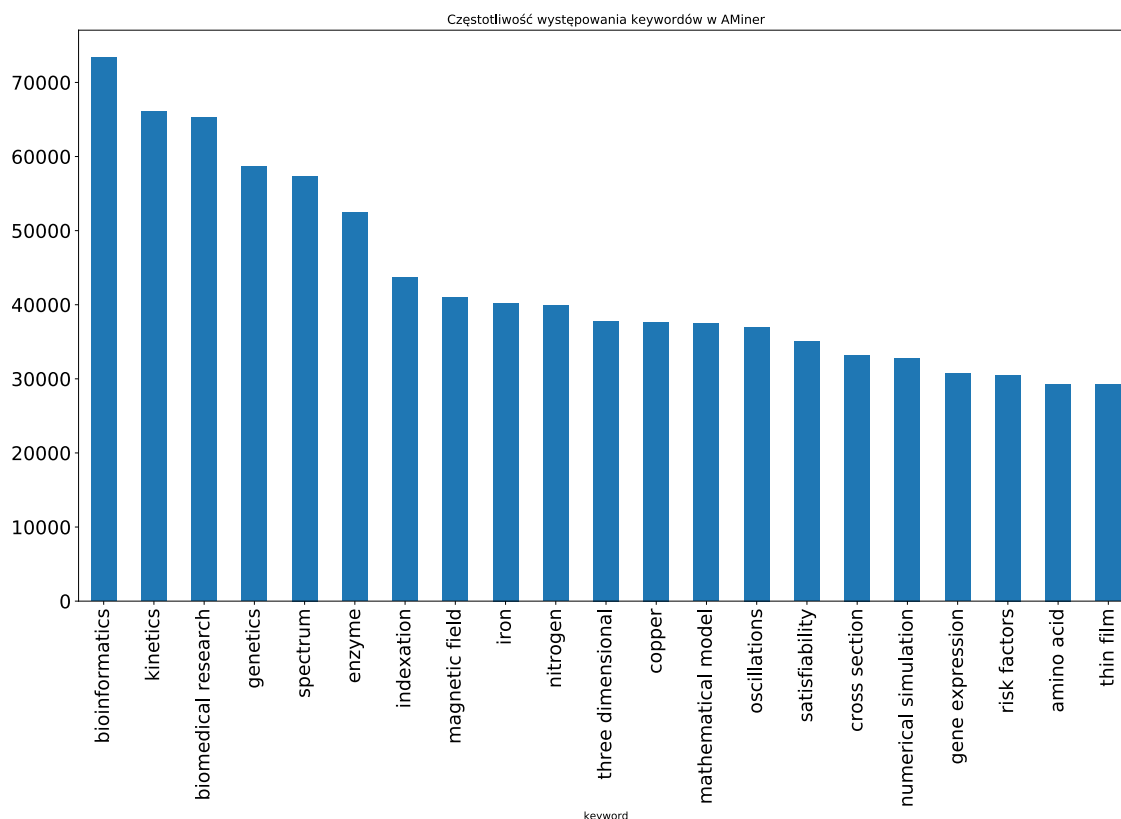
Fakt występowania dla pojedynczej publikacji kilku słów kluczowych nie jest przeszkodą, ponieważ nie powtarzały się one w obrębie pojedynczego rekordu. Występowanie kilku keywordów ma również taką zaletę, że oddaje interdyscyplinarność niektórych publikacji. Dodatkowo, dla części publikacji keywordy występowały w językach innych niż angielski lub w postaci tekstów będących fragmentami abstraktów lub tytułami publikacji. Nie wpływały jednak one znacząco na wyniki ponieważ wystąpiło tylko kilka takich keywordów. Wystąpiło również kilka rekordów, dla których podana data jest błędna np.: 2098 czy 2300.

Ostatecznie otrzymane zostały następujące dane:

- liczba słów kluczy 39,570,260
- liczba różnych słów kluczy 7,222,859
- zakres lat 1800 - 2018

Opis lat:	
średnia	2002.49
Odchylenie standardowe	11.08
min	1800
25%	1999
50%	2006
75%	2009
max	2018

Poniżej przy pomocy bar plot zaprezentowane zostało 20 najaczej występujących słów kluczy.



Wykres 1: Częstość występowania słów kluczowych w AMiner

Na końcu dokumentu załączona została tabela zawierająca wykaz najczęściej występującego słowa kluczowego dla każdego roku. W tabeli zliczone zostały ilości występowania poszczególnych słów kluczowych jako najpopularniejszych (tabela poniżej). Jeśli porówna się jakie słowa kluczowe najczęściej były klasyfikowane jako najpopularniejsze w danych latach to zauważyć można, że wyrazy te są słowami występującymi najczęściej wśród wszystkich publikacji. Interesujące może wydawać się tak częste występowanie bioinformatyki jako najpopularniejszego słowa kluczowego, nawet w XIX w., kiedy jest to dziedzina kojarząca się raczej z czasami współczesnymi i rozwojem komputerów. Faktycznie termin “bioinformatyka” pojawił się po raz pierwszy w 1970 roku. Jego twórcami są Paulien Hogeweg i Ben Hesper. Odnosił się do badania procesów informatycznych w systemach biotycznych i jest ściśle związany z biologią molekularną. Możemy z tego wnioskować, że słowa kluczowe dla przynajmniej części z publikacji dopierane są automatycznie przez pewien klasyfikator, który może wykrywać tematy i powiązywać je z wcześniej wyuczonymi klasami. W takim przypadku, jeśli klasyfikator wytrenowany był na nowszych publikacjach może klasyfikować starsze publikacje zawierające podobne słownictwo bądź podobną tematykę słowami kluczowymi które faktycznie w czasie powstania publikacji jeszcze nie funkcjonowały.

Po 2009 roku widać, że najpopularniejsze słowa kluczowe dotyczące publikacji związane są z tematami powiązanymi z uczeniem maszynowym (ang. *machine learning*, ML). Pokrywałoby się to z ponownym po latach 50-60 boomem na tematy związane z ML i mającym swoje podłoże w nowych możliwościach mocy obliczeniowych.

Słowo kluczowe	Ilość
bioinformatics	93
biomedical research	42
kinetics	27
genetics	9
iron	9
experimental lab study	3
natural history	3
copper	2
general	2
optimization	2
medical education	2
data mining	2
socio economic status	1
vectors	1
electro magnetic	1
arsenic	1
physical optics	1
quality of service	1
chromium	1
nervous system	1
evolution	1
biology	1
chemical analysis	1
infinite series	1
variable stars	1
algorithms	1
nickel	1
dislocations	1
numerical simulation	1

### 3 Analiza danych ze zbioru MAG

W przypadku zbioru MAG największymi trudnościami był fakt występowania ogromnej ilości zbędnych kolumn, oraz niewystępowanie jawnej kolumny ze słowami kluczowymi w strukturze danych. Każdy z tych problemów wymagał innego podejścia.

#### Wczytanie danych

W tej sytuacji efektywnym sposobem wczytywania danych okazało wykonywanie operacji linia po linii do tablicy, z której następnie tworzona jest ramka danych. Ta metoda została wybrana ponieważ omawiany zbiór ma nieregularną strukturę i linie posiadają różną liczbę kolumn. Co za tym idzie, niemożliwe jest użycie standardowych do tego celu funkcji np. `read_csv`. Dodatkowym powodem jest decyzja o wcześniejszym podziale pliku wyjściowego na mniejsze (po 1 mln publikacji). Dlatego pierwszym etapem było opracowywanie danych po 1 mln publikacji, aby następnie połączyć uzyskane tak DataFrame'y o pożądanej strukturze i dokonać właściwej analizy. Pierwszym etapem pracy było usunięcie niepotrzebnych kolumn, oraz wierszy które nie zawierały interesujących nas danych.

## Słowa kluczowe

Aby analizować trendy w nauce na przestrzeni lat najwygodniej analizować występowanie słów kluczowych, wraz z ilością cytowań publikacji, które je zawierają. Aby tego dokonać potrzebna jest wygodna kolumna zawierająca wszystkie słowa klucze przypisane do publikacji. W zbiorze MAG okazało się to być utrudnione, ponieważ nie występowała kolumna **keywords** zawierająca listę słów kluczowych. Występowała za to kolumna **fos**, określana przez Open Academic Graph jako "fields of study", która w strukturze danych wyglądała następująco:

```
"fos": [{ "name": "Computer_network", "w": 0.451811373 }, { "name": "Audio_signal", ...
```

gdzie **name** jest to hasło, które można utożsamiać ze słowem kluczowym, **w** jest wagą tego słowa (w tej analizie pomijane). Czyli to co trzeba było wykonać, to podzielenie kolumny zawierającej **fos** na pojedyncze hasła, oddzielenie od nich i usunięcie części zawierającej wagi, oraz połączenie wszystkich powstałych kolumn w jedną. Cały proces zajmował większość czasu wykonywania się skryptu. Oprócz tego zużywał bardzo dużo pamięci RAM. Z tego powodu niezbędne było opracowywanie danych seriami po ok 0,7 mln publikacji na raz, aby końcową analizę przeprowadzić osobno z gotowymi już DataFrame'ami.

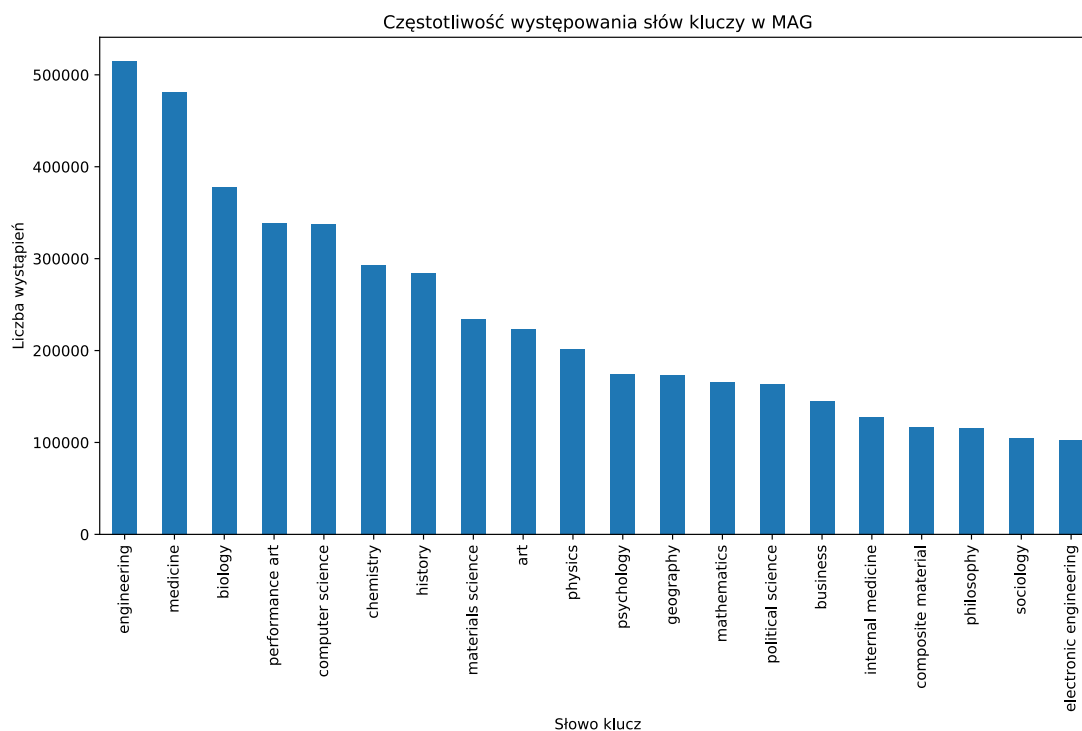
## Właściwa analiza danych

Ostatecznie udało się przygotować DataFrame z dokładnie 4 269 768 publikacjami zawierających słowa kluczowe, ilości cytowań, oraz rok publikacji. Poniżej znajdują się podstawowe informacje o uzyskanym zbiorze:

- liczba publikacji 4 269 768
- liczba różnych słów kluczy: 238 593
- Zakres lat: 1800 - 2019

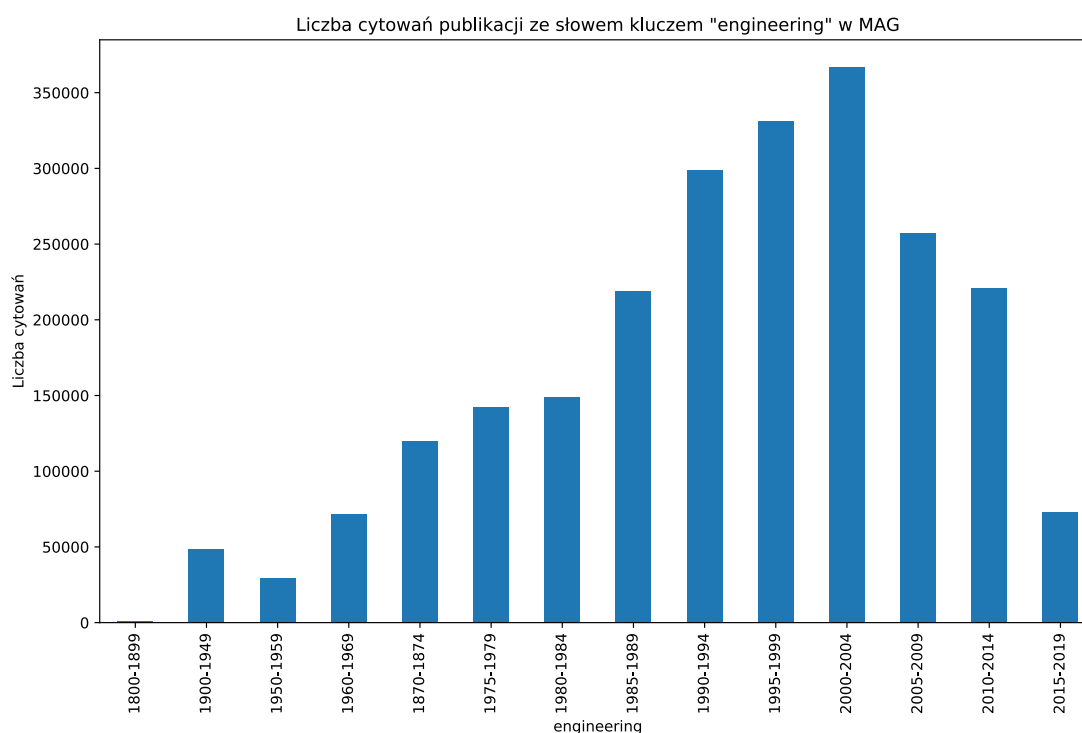
Opis lat:	
średnia	1998,61
Odchylenie standardowe	18,69
min	1800
25%	1992
50%	2005
75%	2011
max	2019

Jako pierwszy został wykreślony histogram 20 najpopularniejszych słów kluczowych na całym zbiorze.



Wykres 2: Częstość występowania słów kluczowych w MAG

Jak widać, najpopularniejszym słowem kluczowym na przestrzeni lat było słowo “engineering”. Jeżeli przeanalizować częstość cytowania tego słowa uzyskamy histogram



Wykres 3: Częstość cytowania publikacji ze słowem kluczowym “engineering” w MAG

Z powyższych wykresów wynika, że najpopularniejsze w analizowanym zbiorze są badania inżynierskie. Można również stwierdzić, że pierwszy rozkwit dziedziny nastąpił na przełomie lat 1900-1949. Powodem może być panowanie w tym czasie dwóch wojen światowych, które znacznie przyspieszyły rozwój technologii. Widać również, że zainteresowanie dziedziną stopniowo rosło aż do wczesnych lat drugiego milenium naszej ery. Następnie można zauważyć spadek zainteresowania, co ciężko jednoznacznie zinterpretować.

Na końcu sprawozdania znajduje się tabela zawierająca dwa najczęściej występujące słowa klucze w publikacjach z każdego roku. Ciekawym sposobem analizy byłoby porównanie tych słów z nagrodą Nobla przyznaną w tym roku w ramach tej dziedziny. Z racji na bardzo częste powtarzanie się takich dziedzin jak np. Medicine; Chemistry, oraz z racji na wydział na którym pisane jest to sprawozdanie ograniczymy się do lat 1965-1969, w których drugim najczęściej powtarzanym słowem kluczowym jest Physics. Poniżej znajduje się tabela zawierająca to porównanie.

Innym powodem zwiększenia popularności tej dyscypliny mogło być przygotowywanie się w tych latach do misji Apollo 11. Jeżeli rozszerzyć zbiór lat od roku 1961 w którym prezydent USA, John F. Kennedy ogłosił, że Amerykanie wylądują na Księżycu przed upływem dekady. Można zauważyć, że najwięcej publikacji wiązało się z: Chemistry, Materials science, oraz Physics. Szczególnie inżynieria materiałowa mocno rezonuje z omawianym wydarzeniem.

Kolejną rzeczą wartą odnotowania jest fakt dużej różnorodności w latach 1800-1900. Wiąże się to najpewniej z małą liczbą publikacji z tych lat w zbiorze. Ciekawe jest jednak to, że znajdują się tam w dużej mierze słowa klucze takie jak: Art; Performance art; Art history, co może wskazywać na dużą rolę (o wiele większą niż dzisiaj) w nauce dziedzin humanistycznych.

rok	laureat	uznanie
1965	Shin'ichirō Tomonaga, Julian Schwinger, Richard Feynman	“Za ich fundamentalne prace z dziedziny elektrodynamiki kwantowej, które wywarły duży wpływ na fizykę cząstek elementarnych”
1966	Alfred Kastler	“Za odkrycie i rozwój optycznych metod badania rezonansu Hertza w atomach”
1967	Hans Bethe	“Za jego wkład do teorii reakcji jądrowych, a w szczególności za odkrycia związane z wytwarzaniem energii w gwiazdach”
1968	Luis Walter Alvarez	“Za decydujący wkład w fizykę cząstek elementarnych, w szczególności za odkrycie wielkiej liczby stanów rezonansowych, co było możliwe dzięki rozwinięciu techniki korzystania z pęcherzykowej komory wodorowej i analizy danych”
1969	Murray Gell-Mann	“Za jego wkład i odkrycia związane z fizyką cząstek elementarnych i ich oddziaływań”

Jak widać prezentowane dane mogą posłużyć do analizy trendów naukowych na przestrzeni lat. Problem z taką analizą niestety wynika z dużej zależności od analizującego. Obiektywne opracowanie tego zbioru jest bardzo trudne, może on jednak posłużyć jako punkt wyjścia do głębszej analizy poszczególnych epok/okresów. Zbiorcza analiza zaprezentowanych 220 lat jest tym bardziej utrudniona, ponieważ występuje dużo niezależnych czynników kształtujących trendy naukowe. Ostatecznie można stwierdzić, że jest to wartościowy zbiór, który mimo problematycznej obróbki jest w stanie pokazywać pewne prawidłowości (szczególnie dla większej ilości danych).



## 4 Wnioski

Uzyskane dane pozwoliły prześledzić trendy w nauce na przestrzeni lat na podstawie słów kluczowych charakteryzujących publikację naukowe publikowane w poszczególnych latach. Po zestawieniu wyników z obu zbiorów widzimy, że tematyka często pokrywa się z dokładnością do ogólniejszych dziedzin. Przykładem tu może być często występująca w zbiorze AMiner bioinformatyka która pokrywa się z medycyną w zbiorze MAG co pozwala przypuszczać, że mamy do czynienia z trendem na publikacje medyczne/biologiczne. Należy jednak pamiętać, że ponieważ niektóre słowa kluczowe z wysokim prawdopodobieństwem przypasowywane są automatycznie przez klasyfikator, należy podchodzić do nich z pewną rezerwą.

Po przeanalizowaniu trendów niektóre z nich jesteśmy w stanie powiązać z konkretnymi wydarzeniami historycznymi czy rozwojem technologiczno-przemysłowym co pozwala potwierdzić słuszność otrzymanych wyników. Niemniej jednak trzeba pamiętać, że przeanalizowana została jedynie część danych ze zbioru OAG, a dodatkowo zbiory MAG oraz AMiner rozwijane są osobno poza OAG. Wyniki otrzymane dla całych zbiorów mogłyby się ostatecznie różnić jednak przeanalizowana część danych i porównanie ich z wcześniej wspomnianymi wydarzeniami historycznymi i rozwojem techniczno-przemysłowym pozwala przewidzieć podobieństwo w wyniku uzyskanym dla całego zbioru i przeanalizowanego fragmentu.

Najczęściej występujące słowa kluczowe na przestrzeni lat w zbiorze AMiner			
Rok	Słowo kluczowe	Rok	Słowo kluczowe
1800	copper	1859	biomedical research
1801	copper	1860	bioinformatics
1803	chemical analysis	1861	biomedical research
1804	physical optics	1862	bioinformatics
1805	variable stars	1863	bioinformatics
1806	chromium	1864	bioinformatics
1808	bioinformatics	1865	biomedical research
1811	bioinformatics	1866	biomedical research
1813	medical education	1867	bioinformatics
1815	iron	1868	bioinformatics
1818	natural history	1869	bioinformatics
1819	infinite series	1870	bioinformatics
1820	natural history	1871	bioinformatics
1821	iron	1872	biomedical research
1822	natural history	1873	bioinformatics
1823	electro magnetic	1874	biomedical research
1824	nervous system	1875	biomedical research
1826	iron	1876	biomedical research
1827	bioinformatics	1877	bioinformatics
1828	dislocations	1878	bioinformatics
1829	bioinformatics	1879	biomedical research
1830	iron	1880	bioinformatics
1831	medical education	1881	biomedical research
1832	biology	1882	bioinformatics
1833	iron	1883	bioinformatics
1834	iron	1884	biomedical research
1835	iron	1885	bioinformatics
1836	nickel	1886	bioinformatics
1837	iron	1887	bioinformatics
1838	arsenic	1888	biomedical research
1839	iron	1889	bioinformatics
1840	biomedical research	1890	bioinformatics
1841	biomedical research	1891	bioinformatics
1842	biomedical research	1892	bioinformatics
1843	biomedical research	1893	bioinformatics
1844	bioinformatics	1894	biomedical research
1845	bioinformatics	1895	biomedical research
1846	biomedical research	1896	biomedical research
1847	bioinformatics	1897	biomedical research
1848	bioinformatics	1898	biomedical research
1849	biomedical research	1899	bioinformatics
1850	biomedical research	1900	biomedical research
1851	bioinformatics	1901	bioinformatics
1852	bioinformatics	1902	bioinformatics
1853	bioinformatics	1903	bioinformatics
1854	bioinformatics	1904	bioinformatics
1855	biomedical research	1905	bioinformatics
1856	bioinformatics	1906	biomedical research
1857	bioinformatics	1907	biomedical research
1858	bioinformatics	1908	bioinformatics

Najczęściej występujące słowa kluczowe na przestrzeni lat w zbiorze AMiner			
Rok	Słowo kluczowe	Rok	Słowo kluczowe
1909	biomedical research	1959	biomedical research
1910	bioinformatics	1960	bioinformatics
1911	bioinformatics	1961	bioinformatics
1912	bioinformatics	1962	bioinformatics
1913	biomedical research	1963	experimental lab study
1914	bioinformatics	1964	experimental lab study
1915	bioinformatics	1965	experimental lab study
1916	bioinformatics	1966	bioinformatics
1917	bioinformatics	1967	bioinformatics
1918	bioinformatics	1968	bioinformatics
1919	bioinformatics	1969	bioinformatics
1920	biomedical research	1970	bioinformatics
1921	biomedical research	1971	bioinformatics
1922	biomedical research	1972	kinetics
1923	bioinformatics	1973	kinetics
1924	biomedical research	1974	kinetics
1925	biomedical research	1975	kinetics
1926	bioinformatics	1976	bioinformatics
1927	bioinformatics	1977	kinetics
1928	bioinformatics	1978	kinetics
1929	bioinformatics	1979	kinetics
1930	bioinformatics	1980	kinetics
1931	bioinformatics	1981	kinetics
1932	bioinformatics	1982	kinetics
1933	bioinformatics	1983	kinetics
1934	biomedical research	1984	kinetics
1935	bioinformatics	1985	kinetics
1936	biomedical research	1986	kinetics
1937	bioinformatics	1987	kinetics
1938	bioinformatics	1988	kinetics
1939	bioinformatics	1989	kinetics
1940	bioinformatics	1990	kinetics
1941	bioinformatics	1991	kinetics
1942	bioinformatics	1992	kinetics
1943	bioinformatics	1993	kinetics
1944	bioinformatics	1994	kinetics
1945	bioinformatics	1995	kinetics
1946	biomedical research	1996	kinetics
1947	biomedical research	1997	kinetics
1948	bioinformatics	1998	kinetics
1949	bioinformatics	1999	kinetics
1950	bioinformatics	2000	genetics
1951	bioinformatics	2001	genetics
1952	bioinformatics	2002	genetics
1953	bioinformatics	2003	genetics
1954	biomedical research	2004	genetics
1955	bioinformatics	2005	genetics
1956	biomedical research	2006	genetics
1957	bioinformatics	2007	genetics
1958	bioinformatics	2008	genetics

Najczęściej występujące słowa kluczowe na przestrzeni lat w zbiorze AMiner			
Rok	Słowo kluczowe	Rok	Słowo kluczowe
2009	data mining	2014	algorithms
2010	bioinformatics	2015	general
2011	numerical simulation	2016	general
2012	optimization	2017	data mining
2013	optimization	2018	vectors

Dwa najczęściej występujące słowa kluczowe na przestrzeni lat w zbiorze MAG					
rok	I słowo kluczowe	II słowo kluczowe	rok	I słowo kluczowe	II słowo kluczowe
1800	Performance art	Art	1850	Art	Performance art
1801	Performance art	Art	1851	Physics	Medicine
1802	Performance art	Art	1852	Physics	Medicine
1803	Performance art	Art	1853	Physics	Art
1804	Art	Art	1854	Medicine	Physics
1805	Art	Performance art	1855	Physics	Performance art
1806	Art	Performance art	1856	Performance art	Art
1807	History	Performance art	1857	Performance art	Medicine
1808	Art	Performance art	1858	Physics	Performance art
1809	Performance art	Art	1859	Performance art	Physics
1810	Performance art	Art	1860	Performance art	Art
1811	Performance art	Art	1861	Performance art	Medicine
1812	Art	Performance art	1862	Performance art	Art
1813	Art history	Performance art	1863	Medicine	Performance art
1814	Performance art	Art history	1864	Physics	Medicine
1815	Performance art	Art history	1865	Physics	Astrophysics
1816	Performance art	Art	1866	Physics	Art
1817	Art	Performance art	1867	Physics	Performance art
1818	Performance art	Art history	1868	Physics	Astrophysics
1819	Art	Art history	1869	Physics	Performance art
1820	Performance art	Art	1870	Performance art	Art
1821	Performance art	Art	1871	Performance art	Art
1822	Art	Performance art	1872	Medicine	Art
1823	Performance art	Art	1873	Medicine	Performance art
1824	Art	Performance art	1874	Medicine	Art
1825	Art	Performance art	1875	Medicine	Performance art
1826	Physics	Art	1876	Medicine	Performance art
1827	Art	Art	1877	Performance art	Art
1828	Performance art	Art	1878	Art	Medicine
1829	Art	Physics	1879	Medicine	Art
1830	Physics	Performance art	1880	Art	Performance art
1831	Performance art	Art	1881	Art	Performance art
1832	Performance art	Medicine	1882	Medicine	Art
1833	Art	Art	1883	Art	Art
1834	Performance art	Art	1884	Art	Performance art
1835	Performance art	Art	1885	Performance art	Medicine
1836	Art	Performance art	1886	Medicine	Performance art
1837	Physics	Performance art	1887	Medicine	Performance art
1838	Art	Performance art	1888	Medicine	Art
1839	Art	Performance art	1889	Art	Performance art
1840	Art history	Performance art	1890	Art	Medicine
1841	Medicine	History	1891	Medicine	Art
1842	Performance art	Medicine	1892	History	Performance art
1843	Medicine	Performance art	1893	History	History
1844	Performance art	Art	1894	Medicine	Performance art
1845	Art	Performance art	1895	Medicine	Performance art
1846	Art	Physics	1896	Medicine	Art
1847	Physics	Performance art	1897	Medicine	Art
1848	Physics	Astrophysics	1898	Medicine	History
1849	Physics	Art	1899	Medicine	Performance art

rok	I słowo kluczowe	II słowo kluczowe	rok	I słowo kluczowe	II słowo kluczowe
1900	Art	Performance art	1950	Biology	Medicine
1901	Medicine	Art	1951	Chemistry	Materials science
1902	Medicine	History	1952	Chemistry	Materials science
1903	Medicine	History	1953	Chemistry	Materials science
1904	History	Medicine	1954	Chemistry	Materials science
1905	Medicine	Physics	1955	Chemistry	Materials science
1906	Medicine	History	1956	Chemistry	Materials science
1907	Medicine	History	1957	Chemistry	Materials science
1908	Medicine	History	1958	Chemistry	Materials science
1909	Medicine	History	1959	Chemistry	Materials science
1910	Medicine	History	1960	Biology	Chemistry
1911	Medicine	History	1961	Chemistry	Materials science
1912	Medicine	History	1962	Chemistry	Materials science
1913	Medicine	History	1963	Chemistry	Materials science
1914	Medicine	Performance art	1964	Chemistry	Materials science
1915	Medicine	History	1965	Chemistry	Physics
1916	History	Medicine	1966	Chemistry	Physics
1917	Medicine	History	1967	Chemistry	Physics
1918	Medicine	Art	1968	Chemistry	Physics
1919	Medicine	Performance art	1969	Chemistry	Physics
1920	Medicine	Biology	1970	Biology	Chemistry
1921	Medicine	Biology	1971	Engineering	Chemistry
1922	Medicine	Biology	1972	Engineering	Chemistry
1923	Medicine	Biology	1973	Engineering	Chemistry
1924	Medicine	Biology	1974	Engineering	Chemistry
1925	Medicine	Biology	1975	Engineering	Chemistry
1926	Medicine	Biology	1976	Engineering	Chemistry
1927	Medicine	Biology	1977	Engineering	Chemistry
1928	Medicine	Biology	1978	Engineering	Medicine
1929	Medicine	Biology	1979	Engineering	Chemistry
1930	Biology	Medicine	1980	Biology	Engineering
1931	Medicine	Biology	1981	Engineering	Medicine
1932	Medicine	Biology	1982	Engineering	Medicine
1933	Medicine	Biology	1983	Engineering	Medicine
1934	Medicine	Biology	1984	Engineering	Medicine
1935	Medicine	Biology	1985	Engineering	Medicine
1936	Medicine	Engineering	1986	Engineering	Medicine
1937	Medicine	Engineering	1987	Engineering	Medicine
1938	Medicine	Engineering	1988	Engineering	Medicine
1939	Engineering	Medicine	1989	Engineering	Medicine
1940	Biology	Medicine	1990	Biology	Medicine
1941	Engineering	Medicine	1991	Engineering	Medicine
1942	Medicine	Engineering	1992	Engineering	Medicine
1943	Medicine	Engineering	1993	Engineering	Medicine
1944	Engineering	Medicine	1994	Engineering	Medicine
1945	Engineering	Medicine	1995	Engineering	Medicine
1946	Engineering	Medicine	1996	Engineering	Medicine
1947	Medicine	Engineering	1997	Medicine	Engineering
1948	Engineering	Medicine	1998	Medicine	Engineering
1949	Engineering	Chemistry	1999	Engineering	Medicine

rok	I słowo kluczowe	II słowo kluczowe	rok	I słowo kluczowe	II słowo kluczowe
2000	Biology	Medicine	2010	Engineering	Medicine
2001	Engineering	Medicine	2011	Engineering	Medicine
2002	Biology	Engineering	2012	Engineering	Medicine
2003	Engineering	Biology	2013	Engineering	Medicine
2004	Engineering	Medicine	2014	Engineering	Medicine
2005	Engineering	Medicine	2015	Medicine	Engineering
2006	Medicine	Engineering	2016	Medicine	Chemistry
2007	Medicine	Engineering	2017	Engineering	Medicine
2008	Medicine	Engineering	2018	Performance art	Art
2009	Medicine	Engineering	2019	Mathematics	Engineering