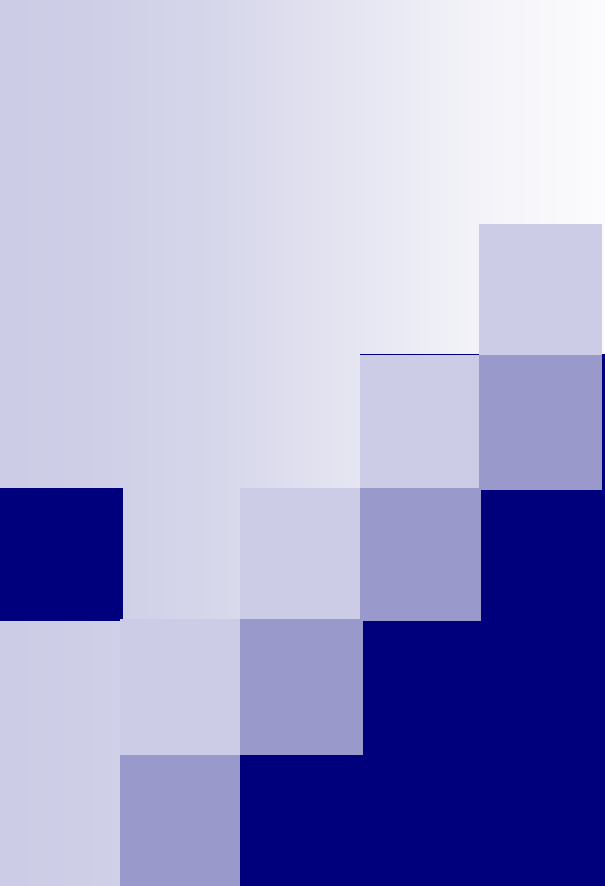


Języki informacyjno- wyszukiwawcze (JIW) oraz język słów kluczowych

Dr hab. Marek Nahotko



Języki informacyjno- wyszukiwawcze (JIW)

Wprowadzenie

Literatura

- Języki informacyjno-wyszukiwawcze. Katalogi rzeczowe / Jadwiga Sadowska, Teresa Turowska. – Warszawa, 1990
- Podstawy budowy tezauryusa / Jadwiga Woźniak-Kasperek. – Warszawa, 2005
- Słowa kluczowe / Wiesław Babik. – Kraków, 2010

Języki informacyjno-wyszukiwawcze (JIW)

- Termin pojawił się w l. 50/60 ub. wieku;
- Wcześniej JIW utożsamiane z zastosowaniami: katalogami systematycznym, działowym, klasowym i przedmiotowym;
- Już w bibliotece Assurbanipala w Niniwie stosowano klasyfikację porządkującą zbiór glinianych tabliczek;
- JIW jest tylko narzędziem, na którym można zbudować katalog rzeczowy.

JIW - definicja

- JIW to język sztuczny o wyspecjalizowanych funkcjach: **odtworzenia treści** i formy dokumentów oraz treści zapytań użytkowników oraz **wyszukiwania** dokumentów w zbiorze informacyjnym w odpowiedzi na zapytanie użytkownika, tzn. których charakterystyki wyszukiwawcze pokrywają się z instrukcjami wyszukiwawczymi (zapytaniami).

Elementy JIW

- Podobnie jak każdy język, w tym także naturalny, każdy JIW posiada:
 - Alfabet – zestaw znaków graficznych używanych do zapisu wyrażeń (słów);
 - Zasób leksykalny (słownictwo) ogół używanych słów (jądro leksyki - stała grupa słów - słownictwo podstawowe);
 - Reguły budowy konstrukcji złożonych (zdań) ze słów (gramatykę).
- Poszczególne JIW różnią się tymi elementami.

Cechy JIW

- Język sztuczny;
- Język o wyspecjalizowanych funkcjach.

Język sztuczny

- Sztuczny, czyli nie naturalny;
- Powstaje przez czyjeś celowe działanie, nie samorzutnie;
- Konstrukcja i słownictwo jest z góry zaplanowane;
- JIW też ulegają zmianom;
- JIW uboższy od naturalnego pod względem budowy, leksyki i funkcji.

Język sztuczny

- Konieczne jest wyeliminowanie zjawisk:
 - Synonimii: słowa o różnym brzmieniu i pisowni, lecz o podobnym znaczeniu (ziemniak, kartofel, pyra),
 - Homonimii: słowa o podobnym brzmieniu i pisowni lecz różnym znaczeniu (zamek),
 - Wieloznaczności: słowo posiada kilka różnych lecz związanych ze sobą znaczeń (żyd, Żyd).

Funkcje j. naturalnego i JIW

- Opisywanie rzeczywistości (f. przedstawieniowa, komunikatywna, opisowa);
- Sygnalizowanie cech psychicznych i fizjologicznych rozmówcy (f. ekspresyjna);
- Oddziaływanie na zachowanie rozmówcy przez tryb rozkazujący, dobór słownictwa (f. impresyjna, imperatywna);
- Podtrzymywanie kontaktu (f. fatyczna);
- Zwracanie uwagi na tekst, jego opis (f. poetycka), treść tekstu (f. metainformacyjna);
- Opisywanie języka jako systemu, jego budowy, funkcji (f. metajęzykowa)
- JIW pełni głównie funkcję przedstawieniową, częściowo funkcję metainformacyjną.

Relacje między wyrażeniami JIW

■ Paradygmatyczne

- relacje między słowami, które uwarunkowane są istnieniem nie językowych, ale logicznych związków między oznaczanymi tymi słowami przedmiotami i zjawiskami
- rodzaje: nadrzędności, podrzędności, równoważności, krzyżowania się, rozłączności, rodzaj-gatunek (mebel-stół), przyczyna-skutek, część-całość (książka-okładka)

Relacje między wyrażeniami JIW

■ Syntagmatyczne

- relacje liniowe, które powstają podczas łączenia słów w grupy i frazy i uwarunkowane są językowo
- współzależności, współrzędności (klasa psy: pudel i kundel)
- kojarzeniowe (samochód-droga)

Typy JIW: języki prekoordynowane

- Jeżeli w JIW indeksowanie dokumentów polega na przyporządkowaniu tych dokumentów do wcześniej już zbudowanych lub tworzeniu w czasie indeksowania wyrażień złożonych, a nazwy tych wyrażień stanowią ChW, wówczas jest to **prekoordynowany JIW**;
- Klasy złożone tworzy się przed i w czasie indeksowania
- Np. klasyfikacje.

Typy JIW: języki postkoordynowane

- Jeżeli w JIW tworzenie wyrażeń złożonych danego JIW z jego wyrażeń elementarnych następuje najpierw podczas indeksowania (potencjalnie) a potem podczas wyszukiwania dokumentów (realnie) wówczas jest to **postkoordynowany JIW**;
- Klasy złożone tworzy się podczas wyszukiwania;
- Języki słów kluczowych i deskryptorowe to języki postkoordynowane.

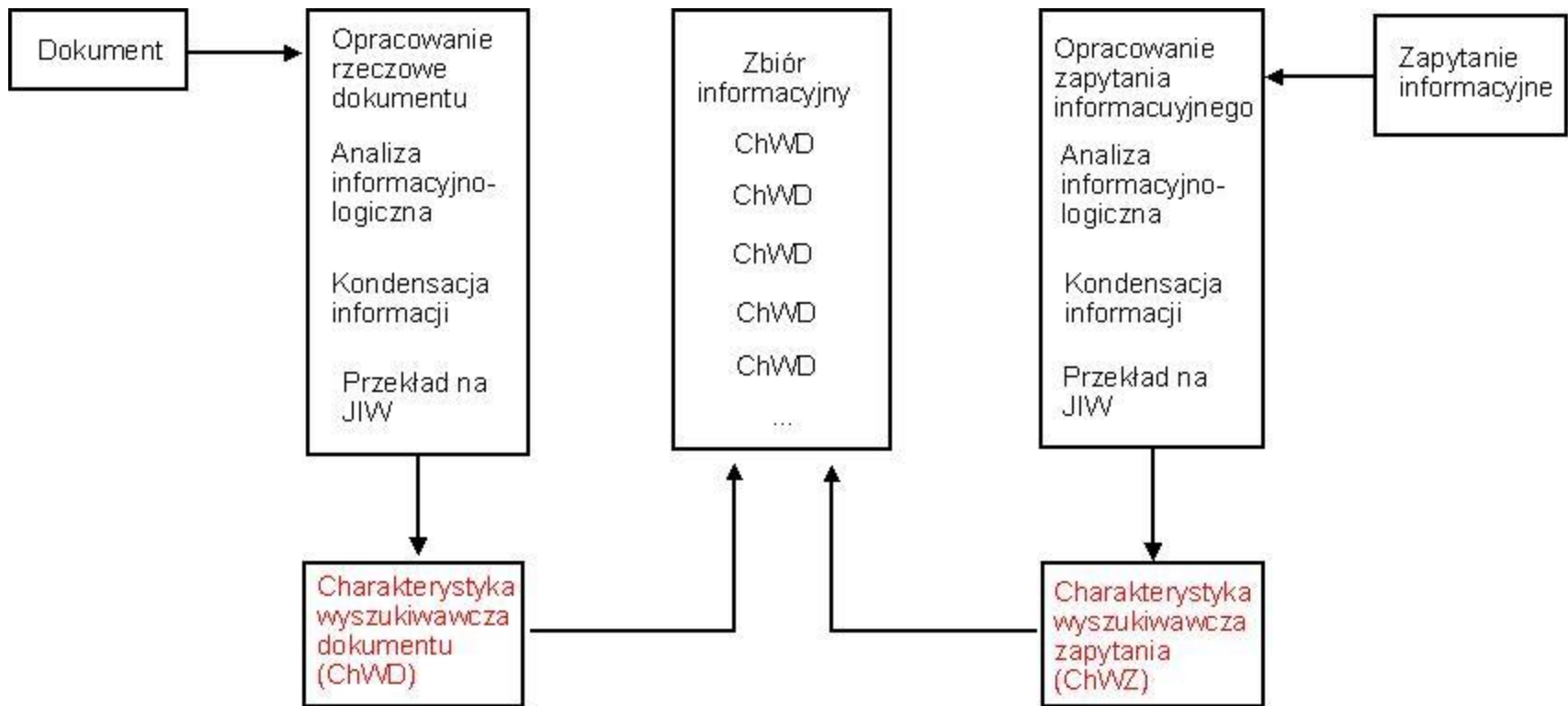
Charakterystyka wyszukiwawcza dokumentu

- Dwie możliwości wyrażania treści dokumentu:
 - Krótki tekst w j. naturalnym (abstrakt, streszczenie, analiza dokumentacyjna);
 - Tekst języka sztucznego – ChWD.

Charakterystyka wyszukiwawcza dokumentu

- Tworzona na podstawie analizy informacyjno-logicznej
- Ustalenie najważniejszych treści dokumentu:
 - Główny temat i przedmiot,
 - Sposób przedstawienia przedmiotu,
 - Przedmioty drugorzędne i poboczne.

JIW w SIW





Języki słów kluczowych

Słowa kluczowe

- **W informatyce:** słowo mające szczególne znaczenie i oznaczające określony nakaz, instrukcję lub deklarację w programie komputerowym;
- **W językoznawstwie:** słowo o ponadprzeciętnej frekwencji w danym tekście;
- **W teorii informacji:** określenie zakresu treści.

Słowa kluczowe – definicja 1

- Wyraz lub wyrażenie wybrane z tytułu lub tekstu dokumentu, charakteryzujące jego treść.

STIN 1979

- Wyrażenie z tekstu dokumentu (często z jego tytułu lub tytułów rozdziałów) lub zapytania informacyjnego charakteryzujące jego treść.

Słownik Encyklopedyczny B. Bojar 2002

Język słów kluczowych

- Dwa znaczenia terminu słowo kluczowe:
 - Każde wyrażenie, wyraz z tytułu lub tekstu dokumentu wykorzystywane do odtwarzania treści lub każde słowo z zapytania;
 - **Jednostka leksykalna języka słów kluczowych.**

Język słów kluczowych – definicja 2

- JIW, którego **słownictwo** tworzą wyrażenia zwane słowami kluczowymi, wybierane z tekstu dokumentu (ew. również spoza niego) lub treści zapytania informacyjnego, nie kontrolowane lub nie w pełni kontrolowane, a **gramatyką** są reguły indeksowania współrzędnego.

Cechy charakterystyczne

- **Forma jednostek leksykalnych** tożsama z formą wyrażen języka naturalnego;
- **Brak kontroli** (lub kontrola częściowa) słownictwa;
- **Słownik uporządkowany** alfabetycznie, tzn. nie uwzględniający żadnych relacji między jednostkami leksykalnymi;
- **Gramatyka** oparta na indeksowaniu współrzędnym (gramatyka zerowa);
- **Postkoordynacja.**

Pochodzenie

- Język słów kluczowych wywodzony od średniowiecza, gdy tworzone katalogi, w których hasłem był główny wyraz tytułu;
- Lata 50. XX w.: indeksy permutacyjne;
- System Unitermów Mortimera Taubego: uniterm – termin jednowyrazowy zapisywany na kartach unitermowych.

MEDYCYNÄ									
0	1	2	3	4	5	6	7	8	9
10 150	11 121 371	2 42	623		55	16 76 126	797	38	9

**TERMINY -
UNITERMY**

INDIE									
0	1	2	3	4	5	6	7	8	9
90	371	142	13	124 244		376	187	428	99

DZIECI									
0	1	2	3	4	5	6	7	8	9
100	31 51 121 371 411		623	244	15	26 46	17	28	19

MEDYCYNÄ. DZIECI. INDIE	371

DOKUMENTY

Funkcje słów kluczowych

■ Funkcje podstawowe

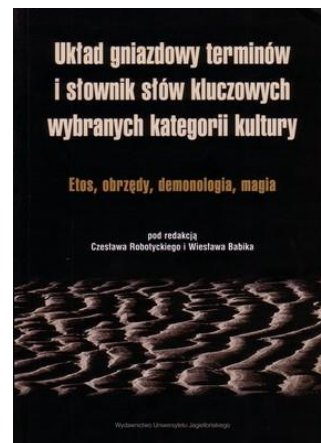
- Funkcja metainformacyjna: odwzorowanie cech dokumentów;
- Funkcja wyszukiwawcza

■ Funkcje dodatkowe

- Funkcja organizująca: porządkowanie (alfabetyczne) zbiorów inf. i dokumentów;
- Funkcja określania zakresów tematycznych zbiorów, dokumentów i informacji.

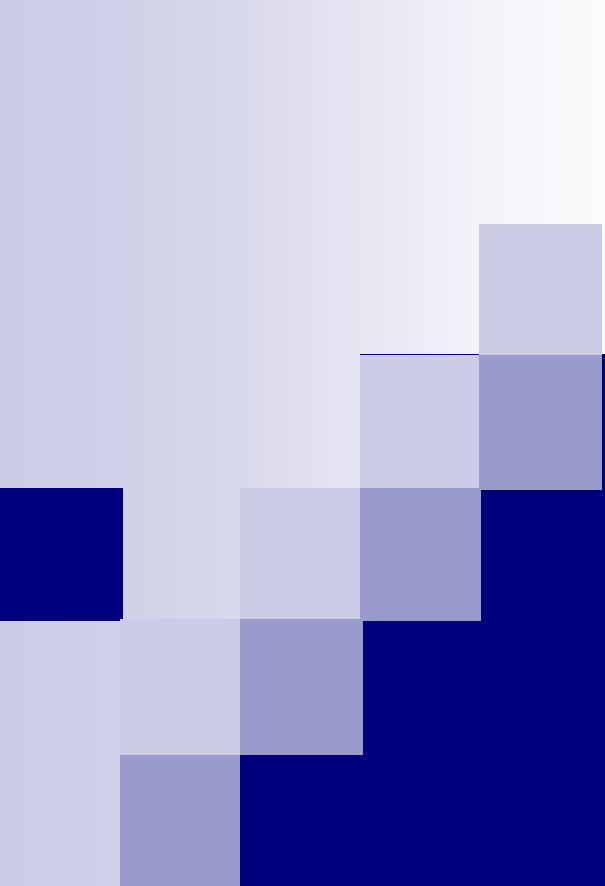
Rodzaje języków słów kluczowych

- Język o słownictwie niekontrolowanym słownikiem (język swobodnych słów kluczowych);
- Język o słownictwie częściowo kontrolowanym (np. wykazy terminów);
- Język o słownictwie kontrolowanym (słownik słów kluczowych).



Zalety i wady

- Łatwość stosowania – nie potrzeba żadnego szkolenia użytkowników;
- Łatwość wprowadzania nowych tematów;
- Brak możliwości wyeliminowania wyrażeń bliskoznacznych i synonimicznych, więc
- Dokumenty o tej samej treści mogą otrzymać różne opisy, w zależności od użytego w nich słownictwa, stąd subiektywizm indeksowania;
- Uzależnienie od języków etnicznych.



Tworzenie słów kluczowych

Tworzenie słów kluczowych

- Z autopsji;
- Podstawa:
 - Język naturalny,
 - Słownik słów kluczowych,
 - Instrukcja indeksowania.

Tworzenie słów kluczowych

- Funkcje słów kluczowych są następujące:
 - Charakterystyka:
 - **Tematu** głównego i ewentualnie tematów peryferyjnych,
 - **Aspektu** w jakim temat jest omówiony.
 - Uzupełnienie lub uściślenie tematu głównego przez podanie dalszych jego cech oraz informacji dodatkowych, np. o miejscu, czasie, punkcie widzenia, rodzaju dokumentu.

Wybór słów kluczowych

- Dwa punkty widzenia:
 - **Kompletność** indeksowania, tzn. odzwierciedlanie przez nie w sposób wyczerpujący (pełny) istotnych elementów treści danego dokumentu z punktu widzenia zakresu tematycznego danego systemu informacyjnego, co jest związane z liczbą słów kluczowych składających się na charakterystykę wyszukiwawczą danego dokumentu;
 - **Szczegółowość** (głębokość) indeksowania, tzn. wprowadzenia do charakterystyki wyszukiwawczej słów kluczowych na takim samym poziomie hierarchii jak użyte w danym dokumencie (np. jeżeli w dokumencie jest mowa o żniwiarkach, należy użyć tego samego określenia jako słowa kluczowego, a nie bardziej ogólnego – maszyny rolnicze).

Gramatyka słów kluczowych

- Rzeczowniki w pierwszym przypadku, w takiej liczbie w jakiej są najczęściej używane, raczej w liczbie pojedynczej niż mnogiej (liczby mnogiej używa się wyjątkowo, tylko w tych przypadkach, gdy dany rzeczownik jest używany wyłącznie w liczbie mnogiej, a użycie liczby pojedynczej byłoby niezręczne lub zmieniałoby treść słowa kluczowego – np. dane techniczne).
- Przymiotniki odnoszące się do danego rzeczownika, w pierwszym przypadku liczby pojedynczej, ewentualnie mnogiej, odpowiedniego rodzaju.

Redakcja słów kluczowych - forma

- Terminy najbardziej rozpowszechnione w literaturze naukowej, technicznej, specjalistycznej, nie potoczne (żargonowe), poprawne językowo i obecnie używane;
- Nazwy pełne, jednostkowe, ogólne, własne, szczegółowe, określonego, specyficznego obiektu, przedmiotu itp.;
- Nazwy proste lub złożone oznaczające przedmioty materialne, procesy, czynności, cechy, właściwości, fakty itp.;
- Skróty, akronimy, gdy są powszechnie znane i stosowane;
- W pewnych przypadkach – gdy jest to konieczne ze względu na treść dokumentu i potrzeby użytkowników – nazwiska, nazwy instytucji, nazwy geograficzne, daty, dane liczbowe, określenie rodzaju dokumentu, cechy formalne dokumentu itp.;
- Przy terminach wieloznacznych należy podać w nawiasach odpowiednie wyjaśnienie, najczęściej określające nazwę dziedziny, której dotyczy dany termin, np. żuraw (budownictwo) żuraw (ornitologia), rak (medycyna), rak (zoologia), rak (taternictwo).

Forma słów kluczowych

■ Terminy:

- Jednowyrazowe, np. technologia;
- Dwuwyrazowe, np. silnik elektryczny;
- Trzywyrazowe, np. elektroniczna technika obliczeniowa;
- Czterowyrazowe, np. metody numeryczne równań nieliniowych.

Liczba słów w terminie

- Słowa kluczowe złożone z kilku wyrazów wprowadza się w następujących przypadkach:
 - gdy znaczenie terminu jako całości nie jest sumą znaczeń poszczególnych składników, lecz stanowi związek o nowym innym znaczeniu, np. **złoty środek, świnka morska, biały kruk, paź królowej**;
 - gdy słowa stanowią nazwy własne, np. **Bolesław Chrobry, Kaplica Zygmuntowska, Czerwone Wierchy**;
 - gdy wyrazy składniowe zachowują swe podstawowe znaczenie, jednak stopień ich zespolenia jest znaczny, np. **skok w dal, silnik spalinowy**.

Forma słów kluczowych

- Słowa kluczowe podaje się zwykle w **pierwszym przypadku liczby pojedynczej**.
- W **liczbie mnogiej** zaleca się podawanie:
 - rzeczowników o charakterze pluralia tantum, np. **nożyczki, drzwi**;
 - nazw roślin, rodów, np. **Radziwiłłowie, Jagiellonowie; storczyki; trawy**;
 - nazw grup związków chemicznych i minerałów, np. **tłuszcze, białka, iły, wapień**;
 - nazw grup osób, np. **dzieci, kombatanci, chłopci, liberałowie, głuchoniemi, Żydzi** (ale nazwy grup zawodowych podaje się w liczbie pojedynczej, np. **górnik, lekarz, nauczyciel**);
 - nazw wyższych grup zwierząt, np. **ssaki, koty, konie** (jako nazwa rodziny, ale **koń arabski** jako nazwa gatunku w liczbie pojedynczej);
 - wyrażień powszechnie stosowanych w liczbie mnogiej, np. **wody przybrzeżne, badania archeologiczne**.

Kolejność słów w terminie

- Na pierwszym miejscu rzeczownik ogólnie charakteryzujący pojęcie, następnie przymiotnik i inne wyrazy stanowiące określenie rzeczownika, uściślające pojęcie ogólne, w kolejności od mniej do bardziej dokładnych;
- Przymiotniki i innego rodzaju określenia powinny zawsze znajdować się bezpośrednio po terminie, do którego się odnoszą (np. samochód osobowy, komunikacja miejska):

RZECZOWNIK → PRZYMIOTNIK OKREŚLAJĄCY TEN
RZECZOWNIK → RZECZOWNIK → DALSZY
PRZYMIOTNIK → itd.

Kolejność słów kluczowych **w** **opisie**

- Kolejność słów kluczowych jest dowolna, występują one jedno za drugim, w sposób nieuporządkowany;
- Traktuje się je jako terminy równorzędne, nawet gdy w rzeczywistości istnieją pomiędzy niektórymi z nich relacje hierarchiczne lub kojarzeniowe.

Interpunkcja

- Znaki stosowane w słowach kluczowych nie są specjalnie „przydzielane”;
- W odróżnieniu od deskryptorów, słowa kluczowe nie zawierają żadnych znaków specjalnych, interpunkcyjnych i alfanumerycznych, mogą być rozdzielane np. przecinkami, kropkami, myślnikami, dwiema spacjami itp.

Liczba słów kluczowych

- Liczba słów kluczowych nie jest limitowana i jest uzależniona jedynie od zakresu i wagi informacji zawartych w indeksowanym dokumencie;
- Liczba ta nie powinna jednak przekraczać średnio 8-15 słów.

Etapy indeksowania

- Zapoznanie się z treścią dokumentu;
- Ustalenie przedmiotu informacji (tego, o czym się mówi) i zakresu informacji (tego, co się mówi o przedmiocie informacji);
- Wybranie i sformułowanie słów kluczowych pochodzących z tekstu dokumentu, a także spoza tekstu dokumentu (słów kluczowych ukrytych), dotyczących zarówno przedmiotu, jak i zakresu informacji;
- Tłumaczenie wybranych słów na jednostki leksykalne języka słów kluczowych;
- Opracowanie leksykograficzne wybranych słów kluczowych według obowiązujących zasad („Instrukcji...”);
- Wykonanie pełnej charakterystyki wyszukiwawczej dokumentu za pomocą wybranych słów kluczowych, zapisanie ich na dokumencie, karcie dokumentacyjnej, w bazie danych itp.;
- Zapisanie słów kluczowych w podręcznym słowniku słów kluczowych.

Przykłady

- Wpływ amoniaku na rozpuszczalność kwaśnych zanieczyszczeń gazowych w wodzie atmosferycznej.
Słowa kluczowe : AMONIAK, CHEMIA ATMOSFERY, DWUTLENEK SIARKI, DWUTLENEK WĘGLA, KWAS AZOTOWY, KWAS SIARKOWY, RÓWNOWAGA GAZ-CIECZ
- Internetowy serwis lingwistyczny.
Słowa kluczowe : INFORMATYKA, JĘZYKOZNAWSTWO, PRZEKŁAD KOMPUTEROWY
- Wybrane usługi telematyczne dla obszarów wiejskich.
Słowa kluczowe : INFORMATYKA, INTERNET, TELEKOMUNIKACJA, TELEMATYKA, WIEŚ

Przykłady

- Controlling jako instrument zarządzania w przedsiębiorstwie budowlanym.

Słowa kluczowe : CONTROLLING, PRZEDSIĘBIORSTWO BUDOWLANE, ZARZĄDZANIE

- Doskonalenie systemu jakości przez ISO 9000 Plus : analiza porównawcza wyników badań przeprowadzonych w Niemczech i w Polsce.

Słowa kluczowe : CERTYFIKATY, ISO 9000 PLUS, JAKOŚĆ, ZARZĄDZANIE

- Edukacja czytelnicza i medialna w szkole.

Słowa kluczowe : CZYTANIE, CZYTELNICTWO, EDUKACJA CZYTELNICZA, EDUKACJA MEDIALNA, MEDIA, SZKOŁA, ŚRODKI MASOWEGO PRZEKAZU