Wiesław Lubaszewski[*], *Marek Gajęcki* [**]

AUTOMATYCZNA EKSTRAKCJA POWIĄZAŃ SEMANTYCZNYCH Z TEKSTU POLSKIEGO

W artykule przedstawiona została metoda automatycznej generacji listy skojarzeniowej dla danego wyrazu. Lista skojarzeniowa jest zbiorem wyrazów sematycznie związanych z wyrazem wejściowym. Podstawą metody jest wnioskowanie statystyczne. Dla generacji listy niezbędne są: słownik fleksyjny, korpus tekstów oraz słownik frekwencyjny. Otrzymane rezultaty mogą posłużyć dla budowy słownika semantycznego.
Słowa kluczowe: *przetwarzanie języka naturalnego, słownik semantyczny, wnioskowanie statystyczne*

AUTOMATIC EXTRACTION OF SEMANTIC ASSOCIATION FROM POLISH TEXT
In the paper we presents a method for automatic construction of an association list for a particular word. Association list is a set of word, each of which is in a semantic relation with the words,to define. To construct the association list we use statistic reasoning algorythm, which works on the base of the Polish inflection dictionary, text corpus, and the quantitative dictionary created for the corpus. Experiments are encouraging enough – one may think that the association can serve as the base for the entry in a semantic dictionary.
Keywords: *Natural Languaqe Processing, semantic dictionary, statistic reasoning*

1. Wstęp

Zalew informacji tekstowej upowszechnianej elektronicznie sprawił, że zintensyfikowano prace nad konstrukcją programów komercyjnych[1)], mogących automatycznie wyszukiwać, klasyfikować, filtrować, nawet streszczać informację przesyłaną, przechowywaną w postaci dokumentów tekstowych. O jakości takich programów decyduje jakość słowników komputerowych oraz jakość algorytmów wyszukujących powiązania semantyczne tekście.

Badania nad tekstem polskim były do niedawna nierealne wobec braku słownika maszynowego polszczyzny. Dziś, gdy dysponujemy odpowiednim słownikiem [3], można podjąć próbę poszukiwania algorytmów rozpoznających powiązania semantyczne w tekście polskim.

[*] Katedra Lingwistyki Komputerowej, Uniwersytet Jagieloński

[**] Katedra Informatyki, Akademia Górniczo-Hutnicza, Kraków, e-mail: mag@agh.edu.pl

[1)] Zob. Inktomi Search, http://www.inktomi.com/products/search/.

Istnieją zasadniczo dwie metody konstrukcji algorytmu wyszukującego powiązania semantyczne w tekście. Pierwsza opiera się na słowniku semantycznym, takim jak np. WordNet [2] i wnioskowaniu dedukcyjnym [1]. Podstawą drugiej metody jest słownik fleksyjny i wnioskowanie statystyczne [1].

W artykule przedstawiono eksperymentalne algorytmy reprezentujące metodę drugą. Celem naszych algorytmów jest konstruowanie na podstawie korpusu tekstów tzw. listy skojarzeniowej dla dowolnego wyrazu podanego na wejściu.

Lista skojarzeniowa to zbiór wyrazów semantycznie związanych wyrazem wejściowym, np. dla wyrazu *flota* wyrazami semantycznie związanymi są: *okręt*, *marynarz*, *admirał*, *port*, *płynąć* itd. Przykładową listę skojarzeniową przedstawia załącznik A.

2. Założenia

Konstrukcja algorytmów opiera się na założeniu, że ludzka wypowiedź odwzorowuje system powiązań semantycznych istniejący umyśle mówiącego [3]. Odpowiednio duża liczba wypowiedzi (korpus tekstów) pozwala skonstruować mechanizm, wykrywający te powiązania.

Podstawą konstrukcji mechanizmu wnioskowania statystycznego wykrywającego powiązania semantyczne pomiędzy wyrazem definiowanym i definiującym jest określenie zależności pomiędzy częstością występowania wyrazu definiującego w korpusie tekstów, a częstością występowania tego samego wyrazu w zdaniach z wyrazem definiowanym.

Zależność tę interpretujemy następująco:

- jeżeli wyraz definiujący, np. *szalupa*, występuje w korpusie tekstów 12 razy, a w zdaniach z wyrazem definiowanym, np. *flota*, 10 razy, to wyraz szalupa jest silnie skojarzony z wyrazem *flota*;
- jeżeli wyraz definiujący, np. *niebo*, występuje w korpusie tekstów 150 razy, a w zdaniach z wyrazem definiowanym, np. *flota*, 15 razy, to wyraz *niebo* jest słabo skojarzony z wyrazem *flota*.

Można przyjąć, że budowa listy skojarzeniowej jest pierwszym etapem tworzenia słownika semantycznego, w którym znaczenie jest reprezentowane przez zbiór relacji zachodzących pomiędzy wyrazem definiowanym a innymi wyrazami, jak np. w przypadku przedstawionego na rysunku 1 opisu znaczenia wyrazu *flota*, gdzie nazwy relacji są wytłuszczone, a listy wyrazów wchodzących w konkretną relację z wyrazem definiowanym podajemy normalną czcionką [4].

---

[2] Fellbaum Ch. (red.): *WordNet. An Electronic Lexical Database*. MIT Press 1998. Zob. też www.cogsci.princeton.edu.

[3] Założenie to potwierdzają eksperymenty prowadzone przez psychologów już w latach 60. ubiegłego wieku, por. Palermo D.S., Jenkins J.J.: *Word Association Norms: Grade School through College*. U. of Minesota Press 1964.

[4] Przykład przedstawia zbiór relacji semantycznych wyróżnionych wyłącznie dla potrzeb naszego eksperymentu.

*** **flota** ***
category: structure
synonymy: armada
similar to: flotylla, eskadra
is part of: marynarka (wojenna)
consists of: admirał, marynarz, okręt
destination: see okręt
action positive: see okręt
negative: rozproszyć

Rys. 1. Przykład definicji wyrazu w słowniku semantycznym

Zaproponowane algorytmy są wynikiem wielu eksperymentów, niektóre z nich przeprowadzono w formie projektów studenckich realizowanych w roku akademickim 2000/2001, można je obejrzeć w Internecie pod adresem: **htttp://winnie.ics.agh.edu.pl**.

3. Narzędzia

Do budowy listy skojarzeniowej algorytmy korzystają z kilku narzędzi:
- słownika fleksyjnego [3],
- korpusu tekstów,
- słownika frekwencyjnego.

Słownik fleksyjny, a więc system rozpoznawania wyrazu w tekście[5], to narzędzie bez którego nie może się obejść żaden algorytm przetwarzania tekstu. Przykład słownika, którego użyliśmy przy konstrukcji naszych algorytmów, przedstawiono na rysunku 2.

=== **flota** ===
description: *rzeczownik, żeński*;
inflection forms: *flota, floty, flocie, flotę, ..., flotach, flotami, floty*;
form description: *Mianownik Liczby Pojedynczej, Dopełniacz Liczby Pojedynczej, Celownik Liczby Pojedynczej, ..., Mianownik Liczby Mnogiej, ..., Wołacz Liczby Mnogiej.*
word category: *wyraz pospolity*

Rys. 2. Przykład jednostki słownika fleksyjnego

Funkcję słownika fleksyjnego pełnił w naszym eksperymencie serwer leksykalny [2], bowiem baza serwera zawiera jednostki słownikowe z kompletną informacją fleksyjną, którą zaczerpnięto ze słownika języka polskiego [3] dostępnego w Internecie. Opisywane algorytmy wykorzystują jedną z podstawowych funkcji realizowanych przez serwer leksykalny, jaką jest rozpoznawanie wyrazów na podstawie ich form tekstowych.

Korpus tekstów to liczący 35 MB w formacie ASCII zbiór informacji prasowych z serwisu agencyjnego PAP z lat 2000–2001.

Słownik frekwencyjny zawiera informację o tym, ile razy dany wyraz (czyli obiekt zdefiniowany na rysunku 2) wystąpił w korpusie tekstów – za wystąpienie wyrazu uznajemy

[5] Por. Lubaszewski W.: *Gramatyka leksykalna w maszynowym słowniku języka polskiego*. Kraków, IJP PAN 1997.

wystąpienie w tekście jednej z przysługujących mu form fleksyjnych. Słownik frekwencyjny został stworzony dla potrzeb niniejszego eksperymentu w oparciu o użyty w eksperymencie korpus tekstów.

4. Algorytm zerowy

Algorytm rozpoczyna swoje działanie od pobrania definiowanego wyrazu – czyli tego, dla którego chcemy utworzyć listę skojarzeniową.

Wszystkie teksty należące do korpusu są kolejno przeglądane i dzielone na zdania. Za zdanie uważa się część tekstu zakończoną kropką, pytajnikiem lub wykrzyknikiem. Do dalszej analizy są brane tylko te zdania, które zawierają definiowany wyraz.

Wyraz za każde wystąpienie w zdaniu otrzymuje 1 punkt. Po zakończeniu analizy korpusu dla każdego wyrazu dzielimy liczbę otrzymanych punktów przez liczbę wystąpień tego wyrazu w korpusie i wynik wyrażamy w procentach. Liczbę wystąpień wyrazu w korpusie udostępnia funkcja `lw ()` na podstawie słownika frekwencyjnego. Otrzymaną listę sortujemy według wyliczonych wartości, a następnie odrzucamy wyrazy, które otrzymały wartość większą niż 50% lub mniejszą niż 2%. Podane progi ograniczające listę są parametrami algorytmu. Konieczność wprowadzenia wartości progowych wynika z prawa Zipfa[6], zaś samą wartość progów ustalono empirycznie.

Algorytm nie uwzględnia odległości pomiędzy wyrazem definiowanym a wyrazami definiującymi[7]. Wyjątkiem od tej zasady były przymiotniki, które były brane pod uwagę wyłącznie wtedy, gdy występowały w bezpośrednim sąsiedztwie rzeczownika.

Wyniki działania algorytmu zerowego (rys. 3) przedstawiono w załączniku A. Wynika z nich, że stosując proste wnioskowanie statystyczne, można osiągnąć interesujące wyniki, jednak na liście znajdują się wyrazy nie mające związku z definiowanym wyrazem, dlatego stworzone zostały kolejne algorytmy.

```
procedure lista0 (wyraz)
begin
  while not koniec_korpusu do
  begin
    wczytaj_zdanie (zdanie)
    if wyraz ∈ zdanie then
      forall w ∈ zdanie do t [w] := t [w] + 1
  end
forall w ∈ t do t [w] := t [w] / lw (w)
sortuj (t)
forall w ∈ t do
    if not (0.02 < t [w] < 0.5) then usuń (t [w])
  return t
end
```

Rys. 3. Algorytm generujący listę „zerową"

[6] Zipf G.K.: *Human Behavior and the Principle of Least Effort.* Cambridge Mass, Addison-Wesley 1949.

[7] Odległość tę uwzględniają niektóre algorytmy studenckie, zob. http://winnie.ics.agh.edu.pl/badania/tm/index.html.

5. Algorytm pierwszy

Algorytm korzysta z procedury generującej listę „zerową". Dla każdego wyrazu, który znalazł się na liście „zerowej", generujemy wtórną listę „zerową". Następnie zliczamy wystąpienia wyrazów na listach wtórnych, tworząc nową listę punktową. Aby uniezależnić otrzymane wartości punktowe od rozmiaru korpusu, dokonujemy ich normalizacji, dzieląc każdą wartość punktową przez wartość maksymalną (rys. 4). Kolejnym krokiem jest posortowanie otrzymanych elementów wg malejącej wartości punktowej i odrzucenie tych, które otrzymały mniej niż 10%.

```
procedure  lista1 ( wyraz)
begin
  L0 := lista0 (wyraz)
  forall  w ∈ L0 do
  begin
     L1 := lista0 (w)
     forall v ∈ L1 do t [v] := t [v] + 1
  end
  max := max (t)
  forall w ∈ t do t [w] := t [w] / max
  sortuj (t)
  forall w ∈ t do
     if t [w] < 0.1 then usuń (t [w])
  return t
end
```

Rys. 4. Algorytm generujący listę „pierwszą"

Tak jak w algorytmie zerowym występujący tu próg 10% jest parametrem algorytmu. Jego zmiana może przyczynić się do polepszenia jakości generowanej listy skojarzeniowej. Listę skojarzeniową wyprodukowaną przez algorytm 1 zawiera załącznik B.

6. Algorytm drugi

Algorytm ten także korzysta z procedury generującej listę „zerową". Dla każdego wyrazu, który znalazł się na liście „zerowej", generujemy wtórną listę „zerową". Następnie każdemu wyrazowi z listy pierwotnej przydzielamy punkty według następującej zasady: wyraz otrzymuje punkt za każdy wyraz na liście wtórnej, którego punktacja według algorytmu zerowego zawarta jest w przedziale pomiędzy 7% a 55%.

Analogicznie jak w algorytmie generującym listę „pierwszą", aby uniezależnić otrzymane wartości punktowe od rozmiaru korpusu, dokonujemy ich normalizacji, dzieląc każdą wartość punktową przez wartość maksymalną. Kolejnym krokiem jest posortowanie otrzymanych elementów wg malejącej wartości punktowej i odrzucenie tych, które otrzymały mniej niż 20%. Występujące w algorytmie progi są parametrami algorytmu.

Listę skojarzeniową wyprodukowaną przez algorytm z rysunku 5 zawiera załącznik C.

```
procedure lista2 (wyraz)
begin
  L0 := lista0 (wyraz)
  forall w ∈ L0 do
  begin
    L1 := lista0 (w)
    forall v ∈ L1 do
      if 0.05 < L1 [v] < 0.55  then t [w] :=  t [w] + 1
end
max := max (t)
forall w ∈ t do t [w] := t [ w ] / max
sortuj (t)
forall w ∈ t do
    if t [w] < 0.2 then usuń (t [w])
  return t
end
```

Rys. 5. Algorytm generujący listę „drugą"

7. Algorytm trzeci

Lista „trzecia" powstaje poprzez usunięcie części wyrazów z listy „zerowej". W tym celu dla każdego wyrazu znajdującego się na pierwotnej liście „zerowej" generowana jest wtórna lista „zerowa". Jeżeli definiowany wyraz nie wystąpi na liście wtórnej, jest on usuwany z listy pierwotnej. Działanie tego algorytmu przedstawiono na rysunku 6.

```
procedure lista3 (wyraz)
begin
  0 := lista0 (wyraz)
  orall w ∈ L0 do
  begin
    L := lista0 (w)
    if not (wyraz ∈ L) then usuń (L0 [w])
  end
  return L0
end
```

Rys. 6. Algorytm generujący listę „trzecią"

Listę skojarzeniową wyprodukowaną przez algorytm z rysunku 6 zawiera załącznik D.

8. Porównanie wyników

Analiza listy zbiorczej przedstawionej w załączniku E pozwala zauważyć, że najlepszą listę produkuje algorytm drugi (kolumna z nagłówkiem A2). Najwyższą punktację uzyskały tu wyrazy: *okręt*, *marynarz*, *marynarka*, *admirał*. Jeśli teraz przyjmiemy za słownikiem *Word-*

Net [7], że flota to „taktyczne ugrupowanie okrętów"[8], to zauważymy, że wyrazy o najwyższej punktacji na liście nie tylko – jak *okręt* – bezpośrednio wiążą listę z przytoczoną definicją pojęcia *flota*, ale też pokazują implikacje semantyczne, jakie produkuje przytoczona definicja pojęcia *flota* – *marynarz* „obsługuje okręt", *admirał* „jest dowódcą jednostki taktycznej", a flota „jest częścią marynarki wojennej".

Analiza listy wyprodukowanej przez algorytm trzeci pokazuje, że może on być stosowany do oceny reprezentatywności korpusu tekstów. Widać bowiem wyraźnie, że wyrazy obecne na listach wyprodukowanych przez wszystkie algorytmy[9] pokazują, iż informacje prasowe na temat floty zdominowała tragedia okrętu podwodnego „Kursk". Możemy zatem powiedzieć, że korpus tekstów użyty w eksperymencie nie jest reprezentatywny dla pojęcia *flota*.

9. Podsumowanie

Przedstawione algorytmy produkują listy skojarzeniowe wystarczająco dobre, by poddać je ręcznej obróbce polegającej na uporządkowaniu listy według relacji semantycznych, co w efekcie przekształci listę w strukturę przedstawioną na rysunku 1.

Jednak wystarczająco dobre, nie oznacza bezbłędne, bowiem jakość generowanych list skojarzeniowych jest zależna od parametrów występujących w algorytmach, a zaproponowane w artykule wartości parametrów są wynikiem przeprowadzenia eksperymentów na stosunkowo niewielkim zestawie słów. Aby optymalnie określić wartości parametrów, należy przeprowadzić więcej prób przy jednoczesnym zwiększeniu rozmiaru analizowanego korpusu tekstów. Innym czynnikiem mającym wpływ na jakość generowanej listy skojarzeniowej jest występująca w języku polskim [7] wieloznaczność form fleksyjnych. Na przykład w liście dla wyrazu *flota* pojawia się błędnie wyraz *bazie*, co wynika z faktu, że forma *bazie* jest jednocześnie formą wyrazu *baza* „ośrodek dyspozycyjny ..." wyrazu *bazia* „kwiatostan roślin z rodziny kotkowatych ...". Jednak problemu wieloznaczności nie da się rozwiązać poprzez modyfikację algorytmów generujących listy skojarzeniowe. Problem wieloznaczności form fleksyjnych poprawnie może rozstrzygać tylko algorytm posługujący się słownikiem semantycznym, wywiedzionym z list skojarzeniowych.

Literatura

[1] Ajdukiewicz K.: *O wnioskowaniu.*[w:] Logika pragmatyczna, Warszawa, PWN 1974
[2] Gajęcki M.: *Serwer leksykalny języka polskiego.* Computer Science, nr 3, 2001
[3] Lubaszewski W., Wróbel H., Gajęcki M., Moskal B., Orzechowska A., Pietras P., Pisarek P., Rokicka T.: *Słownik fleksyjny języka polskiego.* Kraków, Wydawnictwo Prawnicze LexisNexis 2001
[4] Lyons J.: *Semantics.* T.1 T.2. Oxford U. Press 1981
[5] Miller G.A., Johnson-Laird P.N.: *Language and Perception.* Cambridge, Cambridge U. Press 1976
[6] Fellbaum Ch.(red.): *WordNet. An Electronic Lexical Database.* Cambridge, MIT Press 1998
[7] Szymczak M.: *Słownik języka polskiego.* Warszawa, PWN 1978

[8] fleet – a group of warships organized as a tactical unit. Zob. WordNet http://www.cogsci.princeton.edu/cgi-bin/webwn1.7.1?stage=1&word=fleet.

[9] Wyrazy takie oznaczamy symbolem 1.

Załącznik A

Lista skojarzeniowa wyprodukowana przez algorytm zerowy dla wyrazu `flota`.

ID	A0	Wyraz
1007276	50.0	celowniczy
1115569	50.0	złomowanie
1052483	50.0	odpieranie
1000255	30.3	admirał
1000256	30.3	admirał
1033964	25.0	kursowanie
1083026	25.0	sanki
1099973	22.2	wiceadmirał
1048699	21.8	nurek
1102282	20.0	wszczynanie
1048700	18.5	nurek
1037688	17.9	marynarz
1083487	16.6	sejsmograf
1018654	16.6	filtracja
1034039	14.2	kuter
1021039	14.2	głównodowodzący
1093210	13.4	torpeda
1005246	12.5	biomasa
1060296	12.5	plądrowanie
1000933	12.5	alert
1035786	12.5	lotniskowiec
1102602	12.5	wybaczenie
1036487	11.1	łowisko
1093211	11.1	torpedo
1037784	10.0	masowiec
1032402	10.0	krążownik
1019492	10.0	fregata
1054461	9.0	okręt
1112923	8.9	zatonięcie
1018886	8.3	flank
1003294	7.6	basen
1008907	7.6	cielę
1072818	7.5	przerzucanie
1047861	6.6	niewiadome
1093471	6.6	transporter
1071083	6.4	przedział
1088736	6.2	strzelectwo
1077065	5.7	ratownictwo
1018875	5.5	flanka
1101693	4.9	wrak
1048259	4.7	niszczyciel
1048260	4.7	niszczyciel
1075479	4.5	przystanie
1037683	4.2	marynarka
1103320	4.1	wydobywanie

1062174	4.1	podnóże
1037189	4.1	maltretowanie
1014222	4.0	dowódca
1023382	3.5	hobby
1066301	3.5	porucznik
1000845	3.4	akwen
1093474	3.1	transportowiec
1020707	3.1	ginekolog
1086058	3.1	socjologia
1076941	3.1	raki
1004018	2.9	bazia
1002486	2.9	armator
1010552	2.7	czyszczenie
1075478	2.7	przystań
1021016	2.6	głowica
1016670	2.6	eksplozja
1087938	2.5	starodruk
1077543	2.5	rekrutacja
1064108	2.4	poległy
1068914	2.3	pożegnanie
1095663	2.3	ujście
1093473	2.3	transportowiec
1034724	2.2	ląd
1082291	2.2	rybak
1059166	2.1	pęczak
1066759	2.0	postawienie
1060723	2.0	pływak
1060724	2.0	pływak
1082307	2.0	rybołówstwo

Załącznik B

Lista skojarzeniowa wyprodukowana przez algorytm pierwszy dla wyrazu `flota`.

ID	**A1**	**Wyraz**
1048700	68.0	nurek
1048699	65.9	nurek
1000255	63.8	admirał
1000256	63.8	admirał
1054461	59.5	okręt
1037688	53.1	marynarz
1093210	51.0	torpeda
1112923	51.0	zatonięcie
1099973	46.8	wiceadmirał
1034039	44.6	kuter
1019492	38.2	fregata
1032402	38.2	krążownik
1103320	36.1	wydobywanie
1093211	36.1	torpedo
1101693	36.1	wrak

1037189	34.0	maltretowanie
1037683	34.0	marynarka
1083487	34.0	sejsmograf
1060296	31.9	plądrowanie
1018654	31.9	filtracja
1010552	31.9	czyszczenie
1048260	29.7	niszczyciel
1071083	29.7	przedział
1048259	29.7	niszczyciel
1026693	27.6	kadłub
1031105	27.6	kontradmirał
1037346	25.5	manewry
1018995	23.4	flotylla
1003695	23.4	bandera
1090725	23.4	sztorm
1041055	23.4	nabrzeże
1027466	23.4	kapsuła
1035786	21.2	lotniskowiec
1065869	21.2	port
1093559	21.2	trawler
1092154	21.2	tankowiec
1017529	21.2	eskadra
1118007	21.2	żołd

Załącznik C

Lista skojarzeniowa wyprodukowana przez algorytm drugi dla wyrazu `flota`.

ID	A2	Wyraz
1054461	100.0	okręt
1037688	75.0	marynarz
1037683	59.3	marynarka
1048700	56.2	nurek
1071083	50.0	przedział
1000255	46.8	admirał
1000256	46.8	admirał
1048699	43.7	nurek
1014222	43.7	dowódca
1016670	40.6	eksplozja
1112923	40.6	zatonięcie
1101693	37.5	wrak
1064108	34.3	poległy
1099973	28.1	wiceadmirał
1103320	25.0	wydobywanie
1004018	25.0	bazia
1083487	25.0	sejsmograf
1093210	25.0	torpeda
1032402	25.0	krążownik
1048259	21.8	niszczyciel
1048260	21.8	niszczyciel
1019492	21.8	fregata

Załącznik D

Lista skojarzeniowa wyprodukowana przez algorytm trzeci dla wyrazu `flota`.

ID	A3	Wyraz
1048699	1	nurek
1071083	1	przedział
1004018	1	bazia
1064108	1	poległy
1101693	1	wrak
1093210	1	torpeda
1014222	1	dowódca
1003294	1	basen
1112923	1	zatonięcie
1082307	1	rybołówstwo
1054461	1	okręt
1000256	1	admirał
1102602	1	wybaczenie
1082291	1	rybak
1077065	1	ratownictwo
1002486	1	armator
1099973	1	wiceadmirał
1072818	1	przerzucanie
1034039	1	kuter
1000255	1	admirał
1048700	1	nurek
1037683	1	marynarka
1066759	1	postawienie
1037688	1	marynarz
1019492	1	fregata
1016670	1	eksplozja

Załącznik E

Zestawienie wyników działania poszczególnych algorytmów dla wyrazu `flota`.

ID	LW	A0	A1	A2	A0	A1	A2	A3	Wyraz
1112923	28	8.9	51.0	40.6	1	1	1	1	zatonięcie
1019492	15	10.0	38.2	21.8	1	1	1	1	fregata
1000255	33	30.3	63.8	46.8	1	1	1	1	admirał
1037683	141	4.2	34.0	59.3	1	1	1	1	marynarka
1037688	106	17.9	53.1	75.0	1	1	1	1	marynarz
1048699	16	21.8	65.9	43.7	1	1	1	1	nurek
1054461	322	9.0	59.5	100.0	1	1	1	1	okręt
1099973	9	22.2	46.8	28.1	1	1	1	1	wiceadmirał
1093210	26	13.4	51.0	25.0	1	1	1	1	torpeda
1071083	109	6.4	29.7	50.0	1	1	1	1	przedział
1101693	61	4.9	36.1	37.5	1	1	1	1	wrak
1000256	33	30.3	63.8	46.8	1	1	1	1	admirał

1048700	35	18.5	68.0	56.2	1	1	1	1	nurek
1032402	10	10.0	38.2	25.0	1	1	1	0	krążownik
1014222	429	4.0	12.7	43.7	1	0	1	1	dowódca
1034039	14	14.2	44.6	9.3	1	1	0	1	kuter
1016670	152	2.6	17.0	40.6	1	0	1	1	eksplozja
1048259	21	4.7	29.7	21.8	1	1	1	0	niszczyciel
1004018	119	2.9	10.6	25.0	1	0	1	1	bazia
1083487	3	16.6	34.0	25.0	1	1	1	0	sejsmograf
1103320	24	4.1	36.1	25.0	1	1	1	0	wydobywanie
1064108	124	2.4	14.8	34.3	1	0	1	1	poległy
1048260	21	4.7	29.7	21.8	1	1	1	0	niszczyciel
1066759	143	2.0	0.0	15.6	1	0	0	1	postawienie
1093211	9	11.1	36.1	18.7	1	1	0	0	torpedo
1037189	12	4.1	34.0	15.6	1	1	0	0	maltretowanie
1082307	74	2.0	14.8	9.3	1	0	0	1	rybołówstwo
1060296	4	12.5	31.9	9.3	1	1	0	0	plądrowanie
1010552	18	2.7	31.9	12.5	1	1	0	0	czyszczenie
1003294	26	7.6	19.1	15.6	1	0	0	1	basen
1035786	8	12.5	21.2	12.5	1	1	0	0	lotniskowiec
1102602	20	12.5	19.1	6.2	1	0	0	1	wybaczenie
1082291	88	2.2	17.0	18.7	1	0	0	1	rybak
1077065	35	5.7	10.6	3.1	1	0	0	1	ratownictwo
1002486	51	2.9	6.3	12.5	1	0	0	1	armator
1072818	20	7.5	17.0	0.0	1	0	0	1	przerzucanie
1018654	3	16.6	31.9	9.3	1	1	0	0	filtracja