

Проблема визначення тональності

у контент-аналітичних
дослідженнях

Методологія і методи соціологічних досліджень в
Україні: історія та сучасні проблеми. До 70-річчя
Володимира Паніотто



Визначення тональності

Категоризація текстів або їх фрагментів як таких, що містять у собі позитивні або негативні оцінки, інші вираження ставлення комунікатора до явищ або об'єктів, згадуваних у тексті.

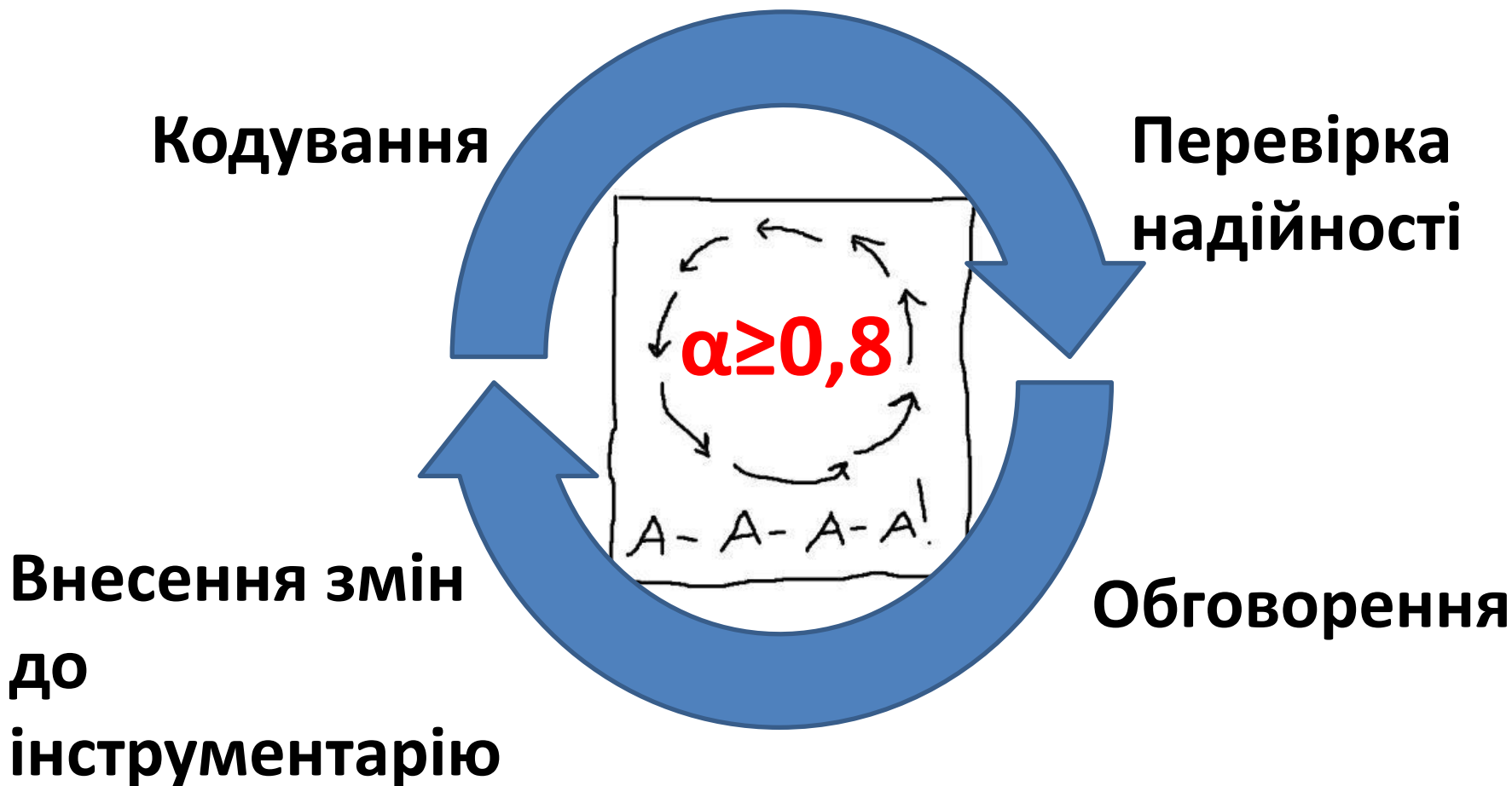


Підходи

- Кодування вручну спеціально навченими кодувальниками.
- Використання словників тональності.
- Машинне навчання.



Кодування вручну





Словниковий аналіз

LOUGHRAN AND MCDONALD FINANCIAL SENTIMENT DICTIONARY

SCALE	NO. OF WORDS	SAMPLE WORDS
Negative	2,337	termination, discontinued, penalties, misconduct, serious, noncompliance, deterioration, felony
Positive	353	achieve, attain, efficient, improve, profitable
Uncertainty	285	approximate, contingency, depend, fluctuate, indefinite, uncertain, variability
Litigiousness	731	claimant, deposition, interlocutory, testimony, tort
Weak Modal Words	27	could, depending, might, possibly
Strong Modal Words	19	always, highest, must, will

Словниковий аналіз

LINIS CROWD

Уважаемый пользователь!

Перед использованием сервиса
Linis Crowd рекомендуем Вам
прочитать инструкцию.

Оцените, пожалуйста, и
положительную, отрица
например: «старик» (0, н
«старикашка» (-2, сильн
из кнопок шкалы с цифр

[Читать полностью](#)

Разметка

Текст для разметки

расчет

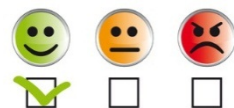
Тональность

Оцените текст:

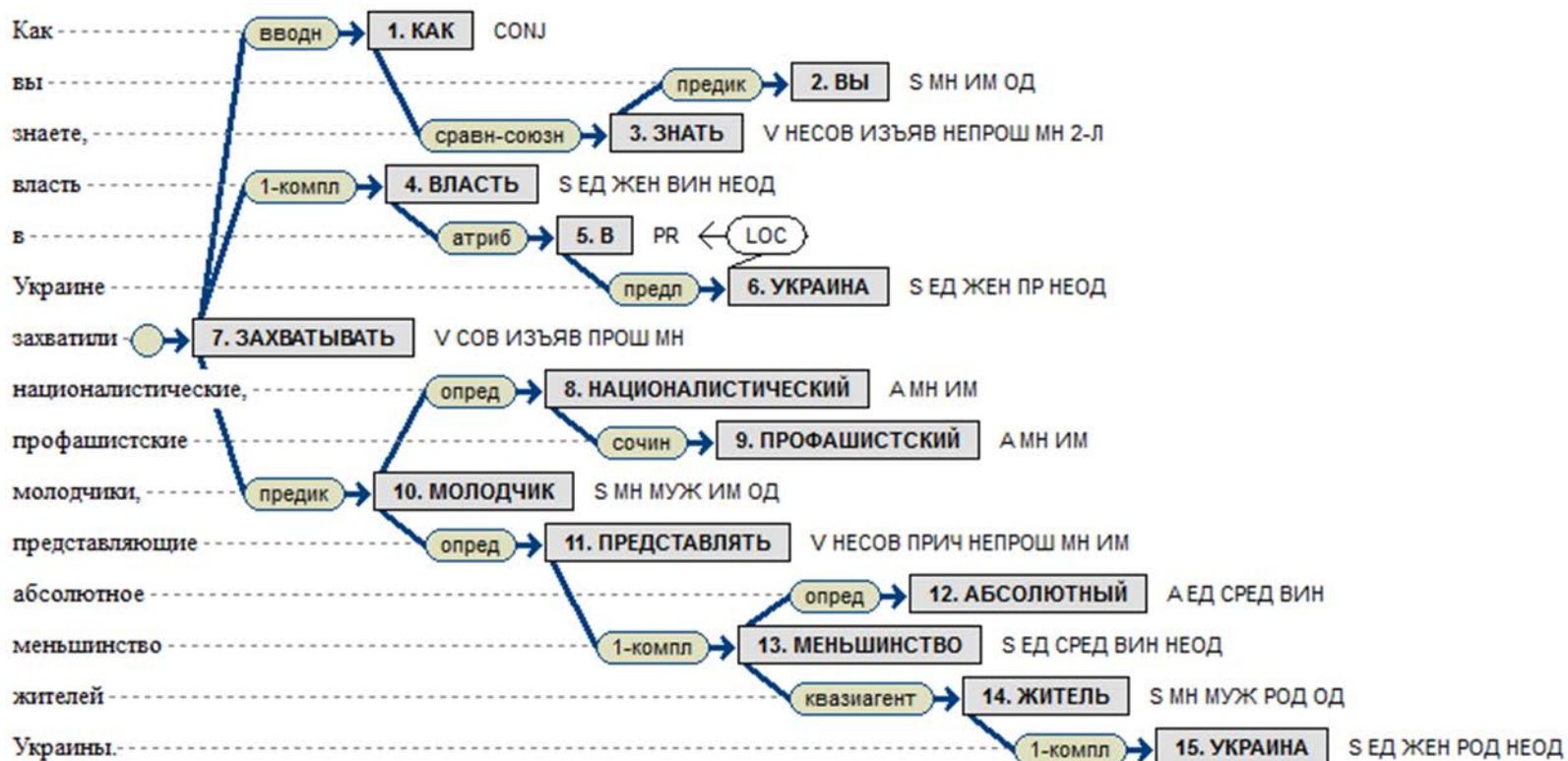
ри этом пункт 54 Постановления
обстоятельств должно быть
условию о цене и недостоверно
незаклученным. Если в
лично я считаю, что Агентство
ему. Не хочет пользоваться
удобство расчетов. Ты -

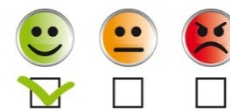
-2	-1	0	1	2
----	----	---	---	---

Words	mean	dispersion	average rate
абориген	-0,25	0,433012702	0
аборт	-1	0,816496581	-1
абрамови	0	0	0
абсолютн	0,333333	0,471404521	0
абстрактн	-0,111111	0,874889764	0
абсурд	-1	0	-1
абсурдно	-1	0	-1
абсурдны	-1	0	-1
абхаз	0	0	0
абхазия	0	0	0
авантюра	-0,33333	0,471404521	0
авантюри	-0,33333	0,471404521	0
аварийны	-1	0,816496581	-1
авария	-1,25	0,433012702	-1
август	0	0	0
авианосе	-0,33333	0,471404521	0
авиацион	0	0	0
авиация	0,111111	0,314269681	0
австралие	-0,33333	0,471404521	0
австралий	0	0	0



Словниковий аналіз





Машинне навчання

```
install.packages((c("RTextTools","tm"),repos
="http://cran.r-project.org")
library(tm)
library(RTextTools)
# Вибір тестових даних (Файл Test.csv)
data <- read.csv2(file.choose(), header = T,
sep = ";")
colnames(data) <- c("Text","Sentiment")
doc_matrix <- create_matrix(data$Text,
language="russian",
removeNumbers=TRUE, stemWords=TRUE,
removeSparseTerms=.998)
container <- create_container(doc_matrix,
data$Sentiment, trainSize=1:90,
testSize=91:187, virgin=FALSE)
RF <- train_model(container,"RF")
RF_CLASSIFY <- classify_model(container,
RF)
```

Text	Sentiment
Населення Росії не має великого бажан	3
А скільки коштує зробити в КМІС За Жи	2
За рейтингами КМІС стирчать вуха Тим	2
КМІС: якби були вибори - у Раду пройш	2
КМІС уточнив, хто має реальні шанси н	2
@euromaydan Це ж КМІС вони весь час	2
"Согласна с автором, ЦИФРЬІ не сходят	2
Навіть на Сході та в Криму приєднувати	3
@UKROPSTEEL схоже, опитування було:	2
"Більшість мешканців Півдня та Сходу н	1
Найкращі серед найгірших: які партії пр	3
Мінськ-3 Це відмова від Євроінтеграції	2
"очень рекомендую к распространени	1
из www.thePO.ST: і не було,і не буде !	3
...	-



Краудсорсінгова розмітка

LINIS CROWD

[Інструкція по використанню](#)

Уважаемый пользователь!

Перед использованием сервиса Linis Crowd рекомендуем Вам прочитать инструкцию.

Оцените, пожалуйста, имеет ли, на Ваш взгляд, предложенное Вам слово положительную, отрицательную или нейтральную эмоциональную окраску, например: «старик» (0, нейтральная), «старец» (+1, умеренно положительная) «старикашка» (-2, сильно отрицательная). Для этого Вам надо нажать на одну из кнопок шкалы с цифрой: -2, -1, 0, +1 или +2.

[Читать полностью](#)

Разметка

Текст для разметки

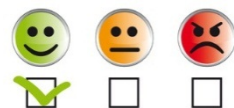
расчет

Тональность

Оцените текст:

-2 -1 0 1 2

ри этом пункт 54 Постановления Пленумов ВС №6 и ВАС №8 устанавливает, что обстоятельств должно быть доказано заинтересованной стороной, а при наличии условия о цене и недостижении сторонами соответствующего соглашения договор не заключенным. Если в договоре это прописать - почему нет? Никаких проблем лично я считаю, что Агент в данном случае - мудак. Так как услуга доставки о нем. Не хочет пользоваться услугой и рисковать еблом - пусть сам катается. Неудобство расчетов. Ты - принципал. Я - агент



«Природна» розмітка

Юрий Уже купил 100% пользователей считают этот отзыв полезным 9 января 2017

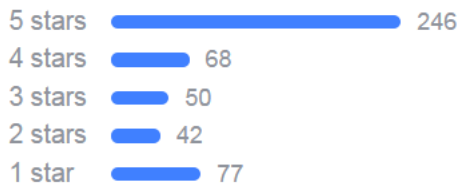
нормальный телефон (бюджетная модель) для работы то что нужно

Достоинства: удобный и простой в эксплуатации

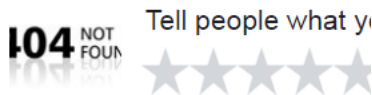
Отзывы о преподавателе

Недостатки: быстро разряжается

3.8 ★ 3.8 of 5 stars
483 reviews



MOST HELPFUL MOST RECE



Larysa Suddya review
December 29, 2016 at 7:

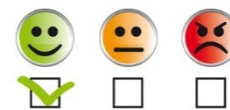
Взагалі то для головного уні
враховуючи економічний за
українцями...шкода. Хотіло:
не маю на увазі дешевий "б
etc. мають набагато доступ
розташовуються на спеціаль
українським брендам, як Lalkarka by Loyanich та ін., відведено малесенький
куточок, дивно дуже виглядає, по жебрацьки. Коротше кажучи, понти і пафос, та
мало раціональності

- + Сдаем у него экзамен. Рассказывал очень интересно, а вот какие баллы известно только ему и ками) 18.05
- + НАЙ-НАЙ-НАЙ краший, від імені релігієзнавців НПУ ім. Драгоманова!!! 20.03
- список литературы ужасный найти практически невозможно 24.07
- никто абсолютный ноль и без палки почитал его литературу -глупости ,аб звание растений в ноосфере и здоровье человека 24.07
- + один из лучших востоковедов Украины, талантливый преподаватель 23.11
- не всегда прав, не знает что такое человеческий фактор 24.09



Кейс: згадування КМІСу в Мережі

1	Text	Sentiment
2	Населення Росії не має великого бажан	3
3	А скільки коштує зробити в КМІС За Жи	2
4	За рейтингами КМІС стирчать вуха Тим	2
5	КМІС: якби були вибори - у Раду пройш	2
6	КМІС уточнив, хто має реальні шанси н	2
7	@euromaydan Це ж КМІС вони весь час	2
8	"Согласна с автором, ЦИФРЫ не сходят	2
9	Навіть на Сході та в Криму приєднувати	3
10	@UKROPSTEEL схоже, опитування було:	2
11	"Більшість мешканців Півдня та Сходу н	1
12	Найкращі серед найгірших: які партії пр	3
13	Мінськ-3 Це відмова від Євроінтеграції	2
181	От інше опитування від КМІС, трохи пр	2
182	"Корисно знати - з урахуванням, що вій	1
183	Вісім запитань лібералам та "дирижист	2
184	70% мешканців Донбасу вірять, що про	3
185	Феноменально: перший номер списку н	2
186	З 18 по 23 травня 2015 року в місті Мик	3
187	18 квітня о 10.00 прес-конференція на т	3
188	С	2



Кейс: згадування КМІСу в Мережі

```
data <- read.csv2(file.choose(), header = T, sep = ";")
colnames(data) <- c("Text", "Sentiment")
doc_matrix <- create_matrix(data$Text,
language="russian", removeNumbers=TRUE,
stemWords=TRUE, removeSparseTerms=.998)
container <- create_container(doc_matrix,
data$Sentiment, trainSize=1:90,
testSize=91:187, virgin=FALSE)
RF <- train_model(container, "RF")
container <- create_container(doc_matrix,
data$Sentiment, testSize=1:187, virgin=T)
RF_CLASSIFY <- classify_model(container, RF)
write.table(RF_CLASSIFY, file.choose(), sep = ";")
```

	FORESTS_LABEL	FORESTS_PROB	
1	3	0.755	
2	2	0.93	
3	2	0.88	
4	2	0.96	
5	2	0.905	
6	2	0.86	
7	2	0.74	
8	3	0.805	
9	2	0.88	
10	2	0.49	
11	3	0.875	
12	2	0.82	
13	1	0.625	

Krippendorff's Alpha Reliability Estimate

	Alpha	LL95%CI	UL95%CI	Units	Observers	Pairs
Nominal	,7002	,6020	,7864	187,0000	2,0000	187,0000