

Ime, priimek:

LV9: Uporabniški poskus in govorne storitve

Na primeru uporabe glasovnih storitev Amazon Aleksa izvedemo poskus z uporabniki, in analiziramo rezultate.

1.1 Uporabniški poskus

Kot uporabniki bomo preskusili govorno storitev dveh sistemov, Amazon Aleksa, in Google Echo.

Primerjali bomo uporabnost obeh sistemov (usability), ter kvaliteto pogovorne storitve.

Izberite si svojo temo pogovora. Od pogovornega sistema želite izvedeti neke uporabne informacije.

Primer:

potujem turistično v London (izberite mesto, državo : USA, VB..).

ogledati si želim galerijo (znamenitosti, nogometno tekmo, koncert ...)

Rezervirati želim hotel, zvedeti želim vremensko napoved, ...

Tematika, cilj vašega pogovora :

1.2 Izvedba uporabniškega poskusa

Izvedite kratek pogovor na vašo tematiko s pogovornim sistemom, pri tem pa spremljajte odziv sistema, skušajte izvedeti iskane informacije.

- pogovor s sistemom A (Aleksa)
- pogovor s sistemom B (Google Echo)

Pri tem bodite tudi pozorni na kvaliteto pogovora (zgovornost sistema, prijaznost, glas ..) ter primerljivost informacij obeh sistemov (koliko vsak sistem poda uporabnih informacij).

1.3 Izpolnite vprašalnike SUS in SUI SQ

Vprašalniki zajemajo dodatne informacije o uporabniku (vpisna št., spol, predhodna uporaba govornih sistemov), ki bodo služile za oblikovanje skupin uporabnikov in interpretacijo.

1. Vprašalnik SUS za sistem A in še enkrat za sistem B

https://docs.google.com/forms/d/e/1FAIpQLSfXUB_vARHK0wb5pB_khmUMzxOa4O2ZHkKBXGH5RZVt6-8C9w/viewform?usp=sf_link

2. Vprašalnik SUIQ-R za sistem A in sistem B

https://docs.google.com/forms/d/e/1FAIpQLSdUGdHCfmvIRxI0MwpcTy-IKqwhTiEWkzqG05y1r9o1p_kUQ/viewform

Appendix 2

See Table 4.

Table 4 The reduced SUI service quality (SUISQ-R) questionnaire

Item	Original	Scale	Item content
1	13	UGO	I would be likely to use this system again
2	17	UGO	I felt confident using this system
3	5	UGO	I could find what I needed without any difficulty
4	1	UGO	The system made me feel like I was in control
5	6	CSB	The system used everyday words
6	11	CSB	The system seemed polite
7	25	CSB	The system seemed professional in its speaking style
8	23	CSB	The system seemed friendly
9	18	SC	The system's voice sounded like a regular person
10	20	SC	The system's voice sounded natural
11	24	SC	The system's voice sounded enthusiastic or full of energy
12	22	V	I felt like I had to wait too long for the system to stop talking so I could respond
13	2	V	The messages were repetitive
14	15	V	The system was too talkative

SUISQ-R scales

User goal orientation (UGO): average items 1–4

Customer service behavior (CSB): average items 5–8

Speech characteristics (SC): average items 9–11

Verbosity (V): average items 12–14 (to reverse score: $V_r = 8 - V$)

Overall: average of UGO, CSB, SC, and V_r

2 Analiza podatkov odgovorov uporabnikov

Raziskovalna hipoteza:

Ali se kvaliteta dvogovora razlikuje med uporabo Aleksa in Google, in za katere faktorje skale?

Faktorji, ki jih izračunamo

- SUS - uporabnost
- UGO (User goal orientation) – uporabnikov cilj
- CSB (Customer service behavior) – obnašanje storitve
- SC (Speech characteristics) – značilnosti govora
- V (verbosity) - gostobesednost

Vstavite analizo podatkov in rezultate.

2.1 Rezultati uporabnosti sistemov SUS

Izračunati moramo rezultat SUS (uporabnost storitve)

How to calculate a SUS score

Looking at a respondent's answers and the corresponding number score for each response, you can tabulate the overall SUS score by using the following framework:

- Add up the total score for all odd-numbered questions, then subtract 5 from the total to get (X).
- Add up the total score for all even-numbered questions, then subtract that total from 25 to get (Y).
- Add up the total score of the new values (X+Y) and multiply by 2.5.

Example scoring:

- Odd = $(4+5+3+4+3) = 19-5 = 14$
Even = $(2+1+3+1+1) = 25-8 = 17$
SUS Score: $(14+17) \times 2.5 = 77.5$
- Odd — questions 1, 3, 5, 7, and 9
Even — questions 2, 4, 6, 8, and 10

By following the scoring tabulation methodology above, you'll then get a SUS score out of 100 (in this case, 77.5/100). Note that this is not a percentage score but a total score out of 100. The average SUS score is 68, and scoring above or below the average will give you immediate insight into the overall usability of the design solution.

Uvozimo podatke iz Excel datoteke v DataFrame

```
import pandas as pd

# Naloži podatke iz lokalne datoteke
fn = 'SUS_01a.xlsx'

df = pd.read_excel(fn)
df
```

Po zgornji definiciji izračunamo stolpce Odd, Even in SUS

```
df['Odd'] = df['Q1']+df['Q3']+df['Q5']+df['Q7']+df['Q9'] - 5
```

The screenshot shows a Jupyter Notebook window with the following content:

Code Cell:

```
df['SUS'] = (df['Odd']+df['Even'])*2.5
df
```

Output:

	T	ID	S	D	A	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Odd	Even	SUS
0	2024-01-03 13:55:20.539	63190140.0	Echo / Alexa	moški	23	4	1	5	1	4	3	5	2	4	1	17	17	85.0
1	2024-01-03 13:55:20.783	NaN	Echo / Alexa	moški	22	1	2	4	2	1	4	4	5	2	1	7	11	45.0
2	2024-01-03 13:59:42.858	63230471.0	Google Voice	ženski	24	3	1	5	1	3	2	4	2	5	1	15	18	82.5
3	2024-01-03 14:01:40.139	63230485.0	Google Voice	moški	24	4	3	3	1	3	4	4	4	5	4	14	9	57.5
4	2024-01-03 14:03:00.615	63190195.0	Google Voice	moški	23	2	2	4	1	2	4	4	3	4	2	11	13	60.0
5	2024-01-03 14:03:13.054	64200087.0	Google Voice	moški	22	2	4	3	1	2	3	4	2	4	1	10	14	60.0
6	2024-01-03 14:04:12.908	63190195.0	Echo / Alexa	moški	23	5	1	4	1	4	3	4	2	4	1	16	17	82.5
7	2024-01-03 14:04:40.664	64200087.0	Echo / Alexa	moški	22	3	2	4	1	4	3	4	2	4	1	14	16	75.0
8	2024-01-03 14:11:20.099	63230485.0	Echo / Alexa	moški	25	4	3	3	3	5	2	4	2	3	4	14	11	62.5
9	2024-01-03 14:11:51.517	63230471.0	Echo / Alexa	ženski	24	5	1	5	3	3	4	4	4	5	4	17	9	65.0
10	2024-01-03 14:13:03.202	63190140.0	Google Voice	moški	23	4	3	5	1	4	2	5	1	4	1	17	17	85.0
11	2024-01-03 14:13:37.285	NaN	Google Voice	moški	22	2	4	4	2	2	4	4	5	1	1	8	9	42.5
12	2024-01-03 14:14:43.625	63190050.0	Google Voice	moški	23	2	3	4	1	4	5	5	2	5	1	15	13	70.0
13	2024-01-03 14:15:36.636	63180095.0	Echo / Alexa	ženski	24	1	3	4	1	2	5	4	2	2	1	8	13	52.5
14	2024-01-03 14:16:02.372	63230493.0	Google Voice	ženski	22	2	2	4	2	2	4	4	2	3	2	10	13	57.5
15	2024-01-03 14:16:56.613	63230493.0	Echo / Alexa	ženski	22	3	2	4	1	3	3	5	2	4	2	14	15	72.5

Code Cell:

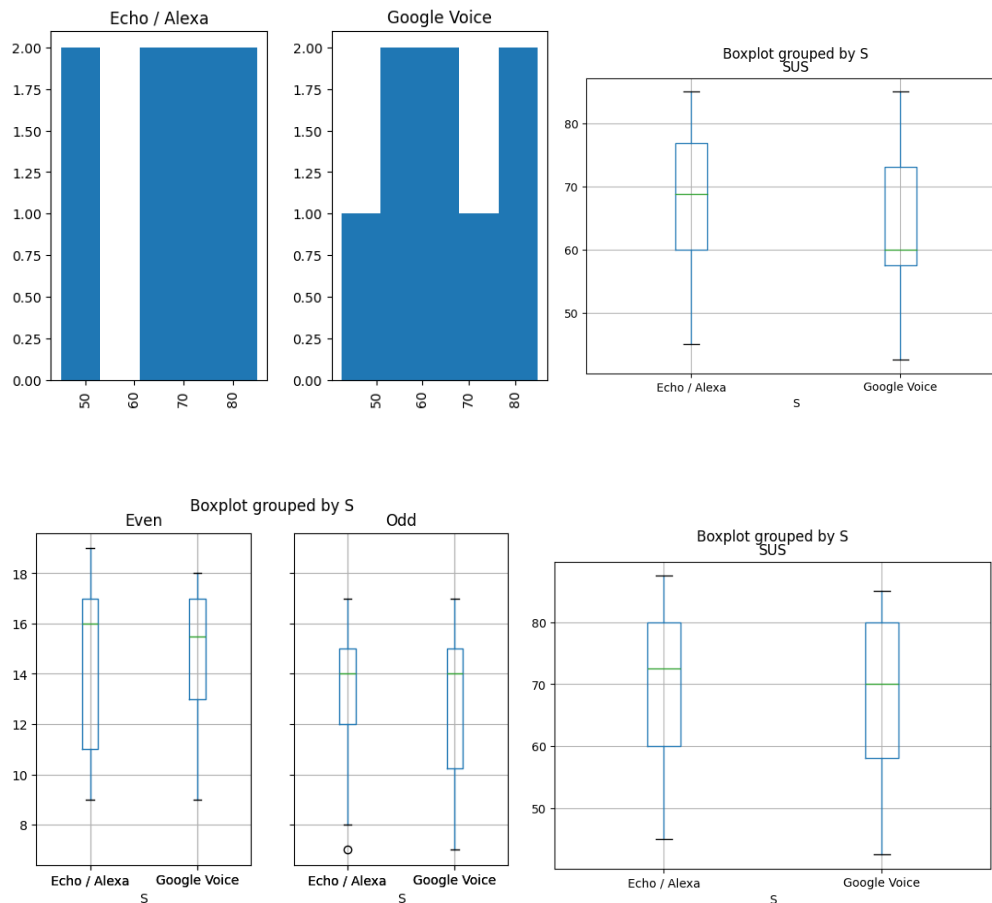
```
# Skupine
S1 = df['S']=='Echo / Alexa'
S2 = df['S']=='Google Voice'
```

Definiramo indekse skupin (Google, Alexa)

```
# Skupine
S1 = df['S']=='Echo / Alexa'
S2 = df['S']=='Google Voice'
```

Izrišemo rezultate po skupinah (boxplot)

```
df.boxplot(by='S', column=['Even','Odd'])  
  
df.boxplot(by='S', column=['SUS'])
```



Preverimo s statističnim testom, ali je porazdelitev rezultatov normalna :

<https://www.statsmodels.org/devel/generated/statsmodels.stats.diagnostic.lilliefors.html>

```
# lilliefors : Ali je porazdelitev normalna, p > 0.05  
from statsmodels.api import stats  
  
S1lil = stats.diagnostic.lilliefors(df[S1]['SUS'])  
print("Normalna porazd p: ", S1lil[1])  
S1lil
```

Vstavi rezultate, ali je normalna porazdelitev.

Interpretacija Rezultatov

· Normalna porazd p: 0.2235

· Rezultat testa: (0.168, 0.223)

· P-vrednost > 0.05 :

predpostavimo, da je porazdelitev normalna.

· P-vrednost ≤ 0.05 :

porazdelitev ni normalna.

Preveri, ali je razlika med sistemi. Izračunaj p-faktor in interpretiraj rezultat.

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html

H_0 : ničelna hipoteza, povprečja za oba sistema so enaka (ni razlike), ($p > 0.05$)

Delovna hipoteza: povprečja se razlikujejo, med sistemi je razlika ($p < 0.05$)

```
# TTest ničelna hipoteza: ni razlike med sistemi,  $p > 0.05$ 
from scipy import stats
stats.ttest_rel(df[S1]['SUS'], df[S2]['SUS'])
```

rezultat, komentar

Normalna porazd p: 0.000999

(0.2953, 0.0009999)

TtestResult(statistic 0.686, pvalue=0.5033897645505774, df=14)

P-vrednost (pvalue): 0.5033897645505774. Ker je p-vrednost večja od 0.05, ni dovolj dokazov da sta si razlilčna za več kot nič. To pomeni, da ne moremo trditi, da obstaja statistično značilna

razlika med povprečji SUS ocen za sistema Echo/Alexa in Google Voice

Statistična vrednost (statistic): 0.6868606776955924

Ta vrednost predstavlja t-statistiko, ki se uporablja za izračun p-vrednosti.

Stopnja svobode (df): 14

Število stopinj svobode, ki se uporablja pri izračunu t-statistike.

2.2 Dodatna naloga: Izriši analize SUS

Uporabi spletno orodje za analizo SUS podatkov sistema A in sistema B:

<https://mixality.de/sus-analysis-toolkit/>

Pripraviti je potrebno csv datoteke za sistema Aleksa in Google in jih naložiti v orodje.

Rezultati, interpretacija:

(Sus rezultat, primerjava z drugimi sistemi, ocena sistema, zanesljivost)

2.3 Faktorji kvalitete pogovora SUI SQ

Izračunaj faktorje:

1. UGO = User goal orientation: uporabnikov cilj
2. CSB = Customer Service Behaviors : obnašanje storitve
3. SC = Speech Characteristics: govorne lastnosti
4. V = Verbosity: gostobesednost, dolgozveznost

V_r : obratni rezultat

Izračunaj po definiciji teh faktorjev:

Rezultat: izračunani faktorji (tabela)

UGO: TtestResult(statistic=-0.4617336564981453,
pvalue=0.6513653305613312, df=14) CSB: TtestResult(statistic=-0.6793883090280631,
pvalue=0.5079695759271889, df=14) SC: TtestResult(statistic=-1.053154786364301,
pvalue=0.31010817331083324, df=14) V: TtestResult(statistic=0.3466578627692474,
pvalue=0.7340035881011885, df=14)

Appendix 2

See Table 4.

Table 4 The reduced SUI service quality (SUISQ-R) questionnaire

Item	Original	Scale	Item content
1	13	UGO	I would be likely to use this system again
2	17	UGO	I felt confident using this system
3	5	UGO	I could find what I needed without any difficulty
4	1	UGO	The system made me feel like I was in control
5	6	CSB	The system used everyday words
6	11	CSB	The system seemed polite
7	25	CSB	The system seemed professional in its speaking style
8	23	CSB	The system seemed friendly
9	18	SC	The system's voice sounded like a regular person
10	20	SC	The system's voice sounded natural
11	24	SC	The system's voice sounded enthusiastic or full of energy
12	22	V	I felt like I had to wait too long for the system to stop talking so I could respond
13	2	V	The messages were repetitive
14	15	V	The system was too talkative

SUISQ-R scales

User goal orientation (UGO): average items 1–4

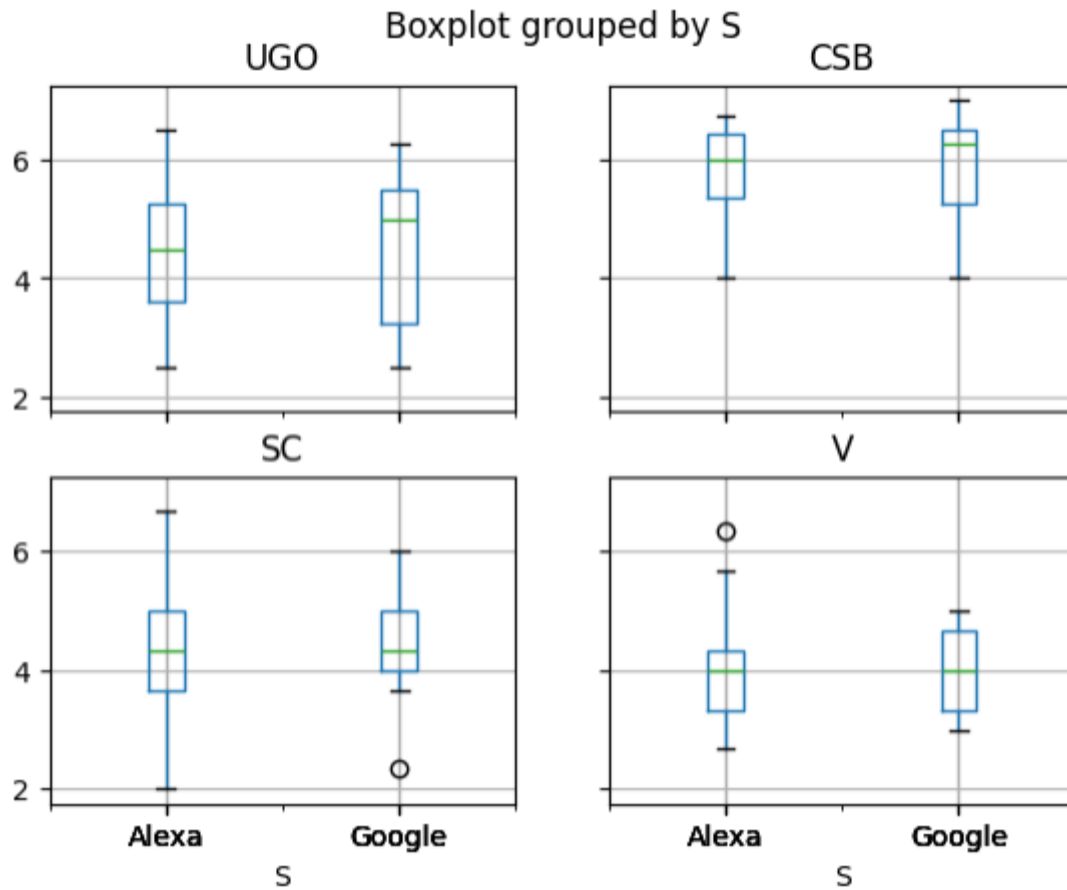
Customer service behavior (CSB): average items 5–8

Speech characteristics (SC): average items 9–11

Verbosity (V): average items 12–14 (to reverse score: $V_r = 8 - V$)

Overall: average of UGO, CSB, SC, and V_r

Boxplot



2.4 Primerjaj sistema Alekxa, Google za 4 faktorje

Statistični T test za posamezne faktorje (ali velja ničelna ali delovna hipoteza):

Parameter UGO: User goal orientation: uporabnikov cilj

T-statistika: -0.4617

P-vrednost: 0.6514

Interpretacija: P-vrednost je večja od 0.05, zato ni statistično značilne razlike v uporabniški usmerjenosti med sistemoma Alexa in Google Voice.

Parameter CSB: Customer Service Behaviors

T-statistika: -0.6794

P-vrednost: 0.5080

Interpretacija: P-vrednost je večja od 0.05, zato ni statistično značilne razlike v obnašanju storitve med sistemoma.

Speech Characteristics (SC):

T-statistika: -1.0532

P-vrednost: 0.3101

Interpretacija: P-vrednost je večja od 0.05, zato ni statistično značilne razlike v govornih lastnostih med sistemoma.

Verbosity (V):

T-statistika: 0.3467

P-vrednost: 0.7340

Interpretacija: P-vrednost je večja od 0.05, zato ni statistično značilne razlike v gostobesednosti med sistemoma.

Boxplot faktorjev po sistemih (Aleksa, Google)