

# SVILUPPATORI E PIATTAFORME CLOUD CONVERSAZIONALI: UTILIZZO E DIFFICOLTÀ RISCONTRATE TRAMITE L'ANALISI DI STACKOVERFLOW

**Candidato:** Cacciamali Matteo

**Relatore:** Luigi De Russis



# COS'È STACKOVERFLOW ?



- Fondata nel 2008, StackOverflow è la community online più grande e affidabile per gli sviluppatori che desiderano apprendere e condividere le proprie conoscenze
- Oltre 50 milioni di programmatori professionisti e aspiranti visitano StackOverflow ogni mese per aiutare a risolvere problemi di codifica e sviluppare nuove competenze

stackoverflow

Search...

Home

PUBLIC

Stack Overflow

Tags

Users

Jobs

Teams

Q&A for work

Learn More

Top Questions

Ask Question

Interesting 352 Featured Hot Week Month

0 votes

0 answers

1 view

Struggle with getting ID from user

java arrays string methods

asked 1 min ago Yashiru99 8

0 votes

0 answers

2 views

Getting unexpected TypeErrorwhen trying to find element on page by text

python python-3.x selenium selenium-webdriver selenium-chromedriver

asked 1 min ago Gregory 130

0 votes

0 answers

10 views

Mysterious behavior in flex-direction: column

css flexbox

modified 1 min ago j08691 165k

0 votes

0 answers

5 views

Madcoda youtube get all channels name

php

modified 1 min ago Alex Howansky 34.7k

0 votes

0 answers

2 views

rxjs Merge Observable and Subject

redux rxjs redux-observable

asked 1 min ago Matthew Crews 1,462

0 votes

1 answer

9 views

Send mail using mailtrap in laravel

php laravel phpmailer mailer

answered 1 min ago FGDeveloper 409

# COS'È DIALOGFLOW ?

Book a flight  
from Los Angeles  
to Hawaii for less  
than \$300

You got it!

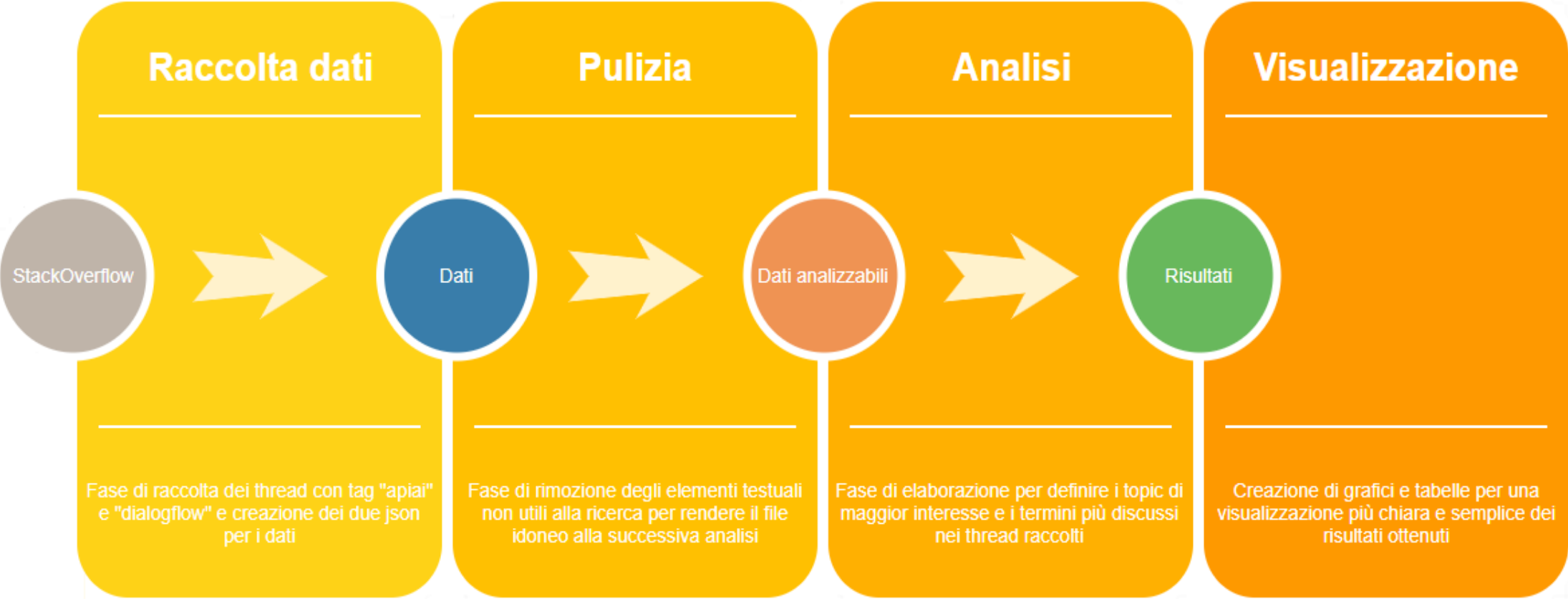


Hi, how can I help?



"Alexa, turn on  
Welcome Home"

"Alexa, turn on  
the TV"



- Nato nel 2010
- Consiste in 173 Q&A communities tra cui StackOverflow
- Fornisce API
- Permetter la creazione di custom filters

# StackExchange





```
"tags": [
  "json",
  "android-intent",
  "drag-and-drop",
  "api.ai"
],
"comments": [
  {
    "comment_id": 77910064,
    "body": "<code>postman</code> is the best tool I have ever seen"
  },
  {
    "comment_id": 77910436,
    "body": "thank you, will give it a try! @sudhakar"
  }
],
"answers": [
  {
    "comment_count": 0,
    "is_accepted": false,
    "answer_id": 45303999,
    "body":
      "<p>i dont know it would help you or not but try postman
      not entirely for the situation you mentioned but removes
      a lot of pain from your head <a
      href=\\\"https://www.getpostman.com/docs/postman/collection
      s/creating_collections\\\" rel=\\\"nofollow
      norereferrer\\\">https://www.getpostman.com/docs/postman/coll
      ections/creating_collections</a></p>\\n"
  }
],
"comment_count": 2,
"is_answered": false,
"answer_count": 1,
"creation_date": 1500986979,
"question_id": 45303858,

"title":
  "Is there any way to create a tool with drag and drop
  mechanism to create a JSON file of intents in api.ai?",

"body":
  "<p>I am looking for a tool or a way through which I can
  create a drag and drop mechanism containing intents
  messages for user inputs and the respective response to
  it. It should then be able to create a JSON file of those
  multiple intents so that I can upload it to api.ai, and
  my intents are then visible there.</p>\\n"
```

Quantificazione dati di StackOverflow	
Data thread più vecchio	Domenica 1 ° gennaio 2017
Data thread più recente	Mercoledì 17 ottobre 2018
Domande	1736
Risposte	1637
Commenti a risposte	2001
Commenti a domande	2075
Media di commenti per risposta	1.2
Media di commenti per domanda	1.2
Domande marcate come chiuse	887
Percentuale di domande chiuse	51.09%

- In data 10 Ottobre 2017 cambia nome



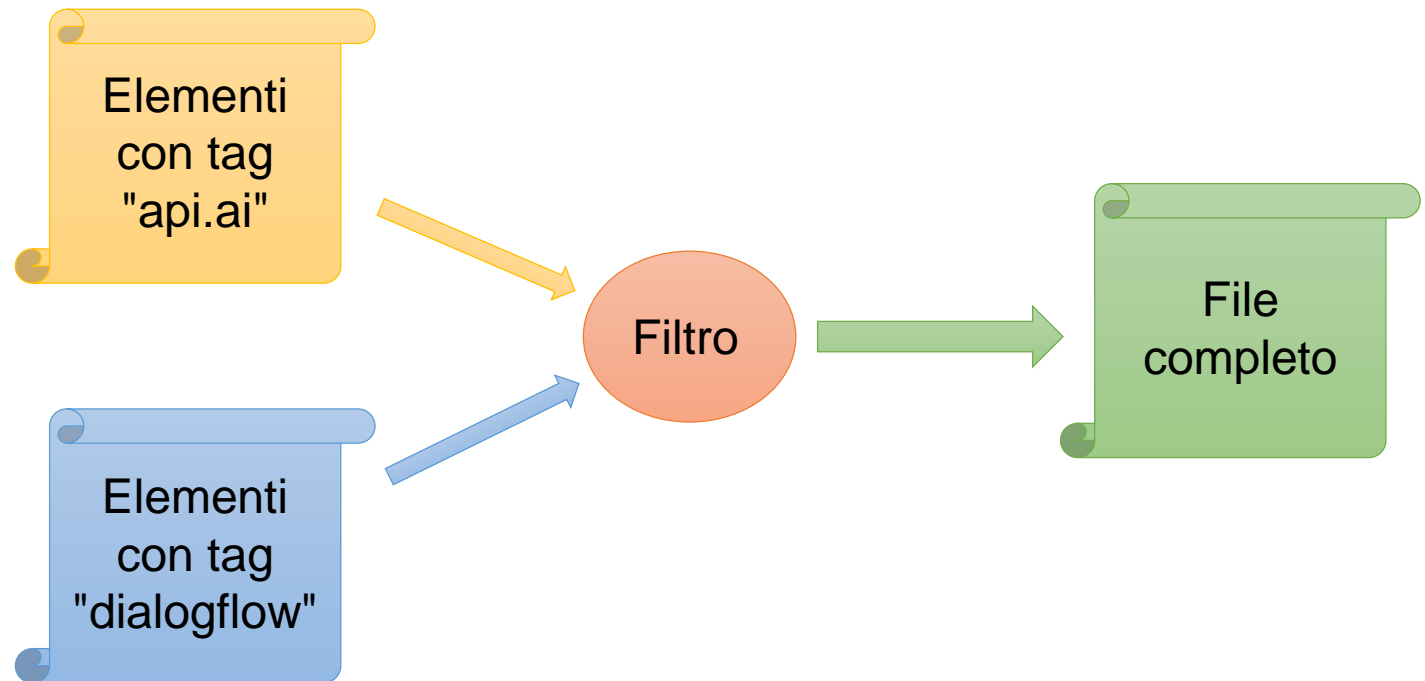
- Aggiunge 2 nuove funzionalità

1. In-line code editor:

ora puoi scrivere la logica di realizzazione, testare e implementare un webhook funzionale direttamente nella console.

2. Multi-lingual agent support:

la creazione di più lingue è ora più semplice che mai. Ora puoi aggiungere altre lingue e impostazioni locali al tuo agente nuovo o esistente.





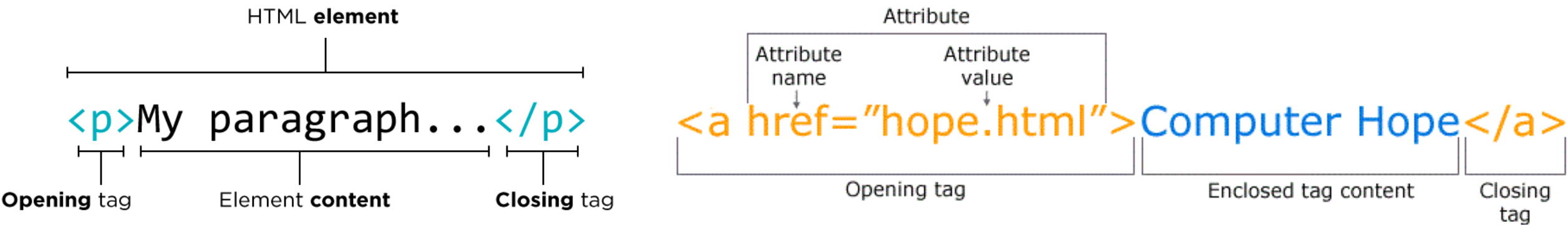
*<p> Now that <a href=\https://developers.google.com/actions/localization/\ rel=\"nofollow noreferrer\">localization is available</a> for Actions on Google, I'm wondering about minimizing latency for user responses. In a stack where an app is using Dialogflow for NLP and then a Google Cloud Function behind that for webhook fulfillment, is it possible to set up cloud functions where a specific localized experience has been established in Actions on Google? In other words, can we set up English-US responses to be provided by a cloud function hosted in us-central1, and English-GB via europe-west1 Google datacenters?</p>\n\n<p>With Alexa & ASK you can set up a language with 'geographical region endpoints' in N America, Europe and India and have Lambda functions fulfill intents from users in those regions.</p>\n\n<p>Dialogflow may not be available in multiple GCP regions, but if it is it would be nice to keep fulfillment local with google cloud functions close to where the user is - if it minimizes conversation delays. Maybe someone has data to suggest it doesn't matter and is fast enough either way.</p>\n"*



re — Regular expression operations

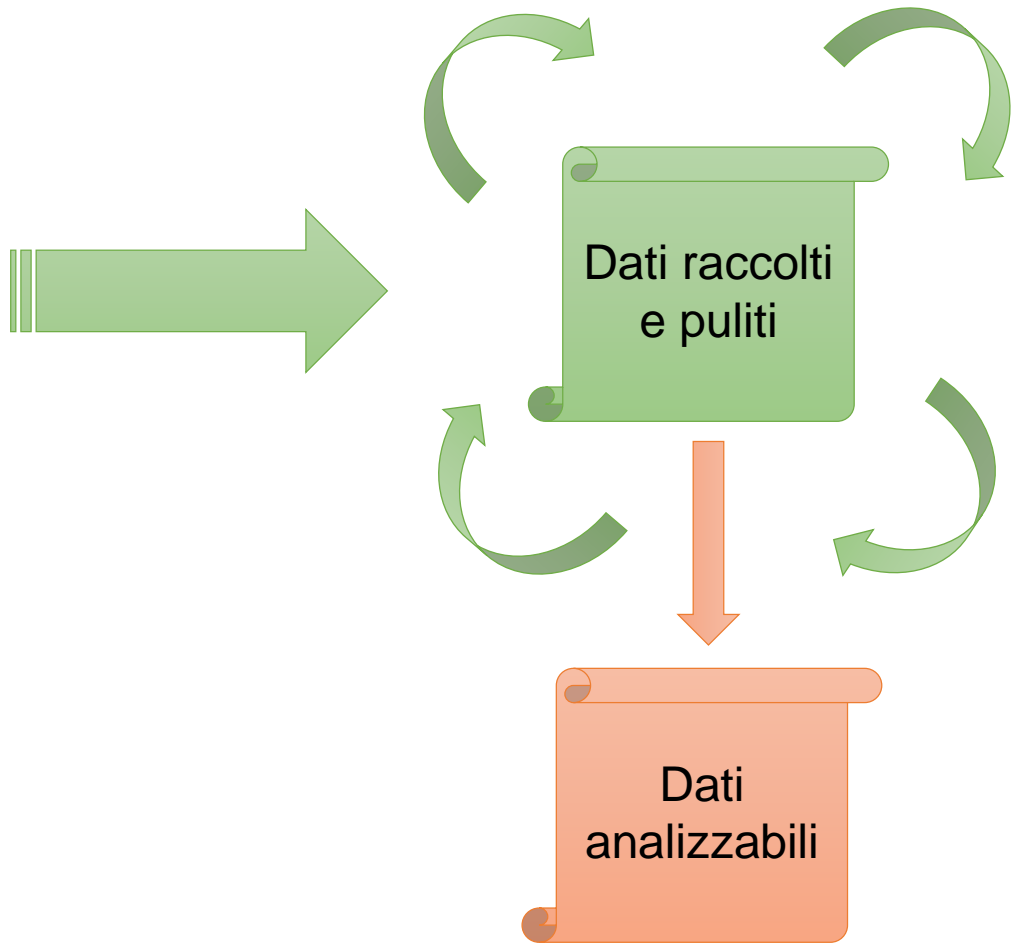
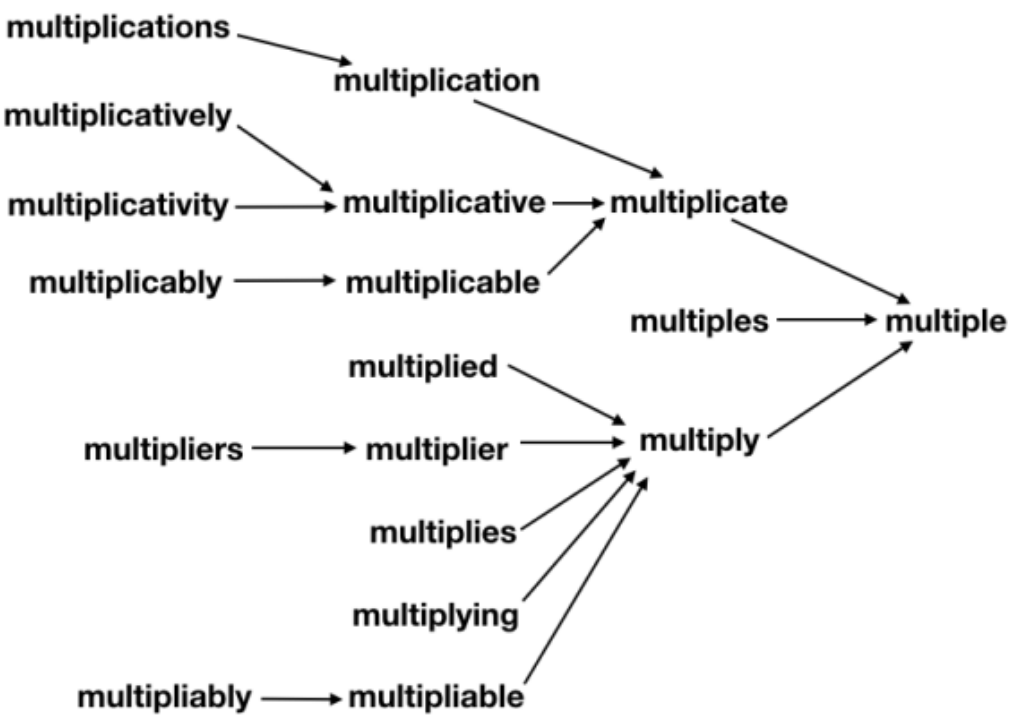
*action wonder minimize latency user response stack app nlp cloud function webhook fulfillment set cloud function specific localize experience establish action set english response provide cloud function host central english gb europe west datacenter alexa amp set language geographical region endpoint america europe india lambda function fulfill intent user region multiple gcp region nice keep fulfillment local cloud function close user minimize conversation delay data matter fast*





- 1. Inserimento di spazi vuoti a inizio e fine testo
- 2. Rimozione dei “ritorno a capo”.
- 3. Conversione di ogni carattere in formato minuscolo
- 4. Rimozione dei tag HTML.
- 5. Rimozione di caratteri speciali HTML ad esempio “&#39;” => apostrofo
- 6. Eliminazione dei numeri
- 7. Rimozione della punteggiatura
- 8. Sostituzione di tutti gli eventuali spazi consecutivi con uno spazio singolo
- 9. Rimozione delle English stop words
- 10. Rimozione degli spazi vuoti a inizio e fine testo inseriti nel passo 1.

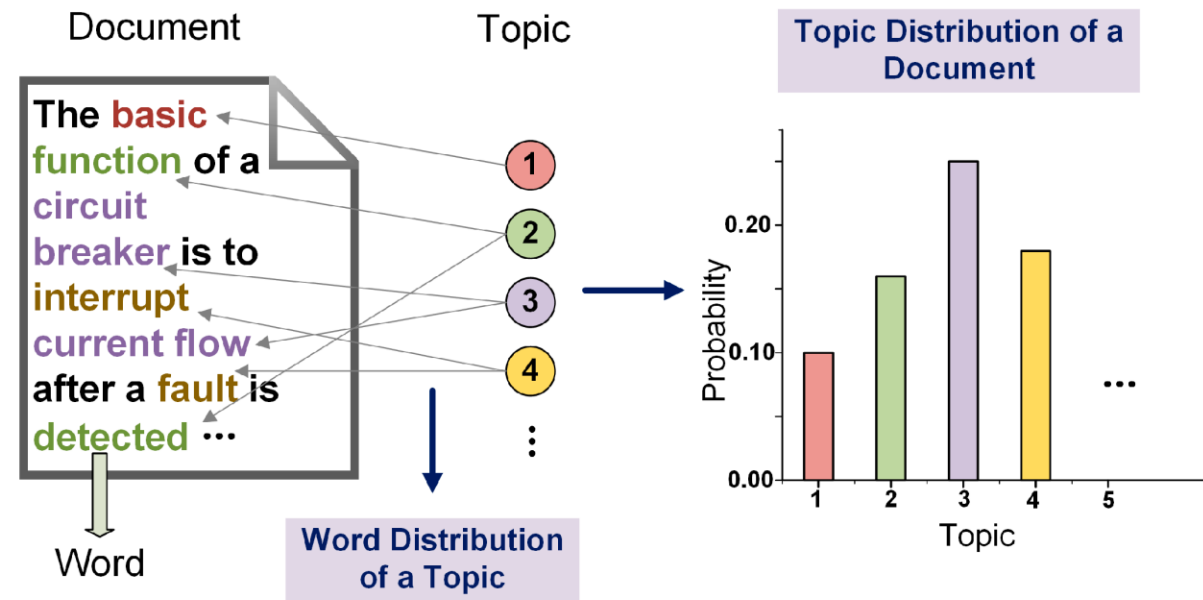
Default English stopwords list	
This list is used in our <a href="#">Page Analyzer</a> and <a href="#">Article Analyzer</a> for English text, when you let it use the default stopwords list.	
a	ourselves
about	out
above	over
after	own
again	same
against	shan't
all	she
am	she'd
an	she'll
and	she's
any	should



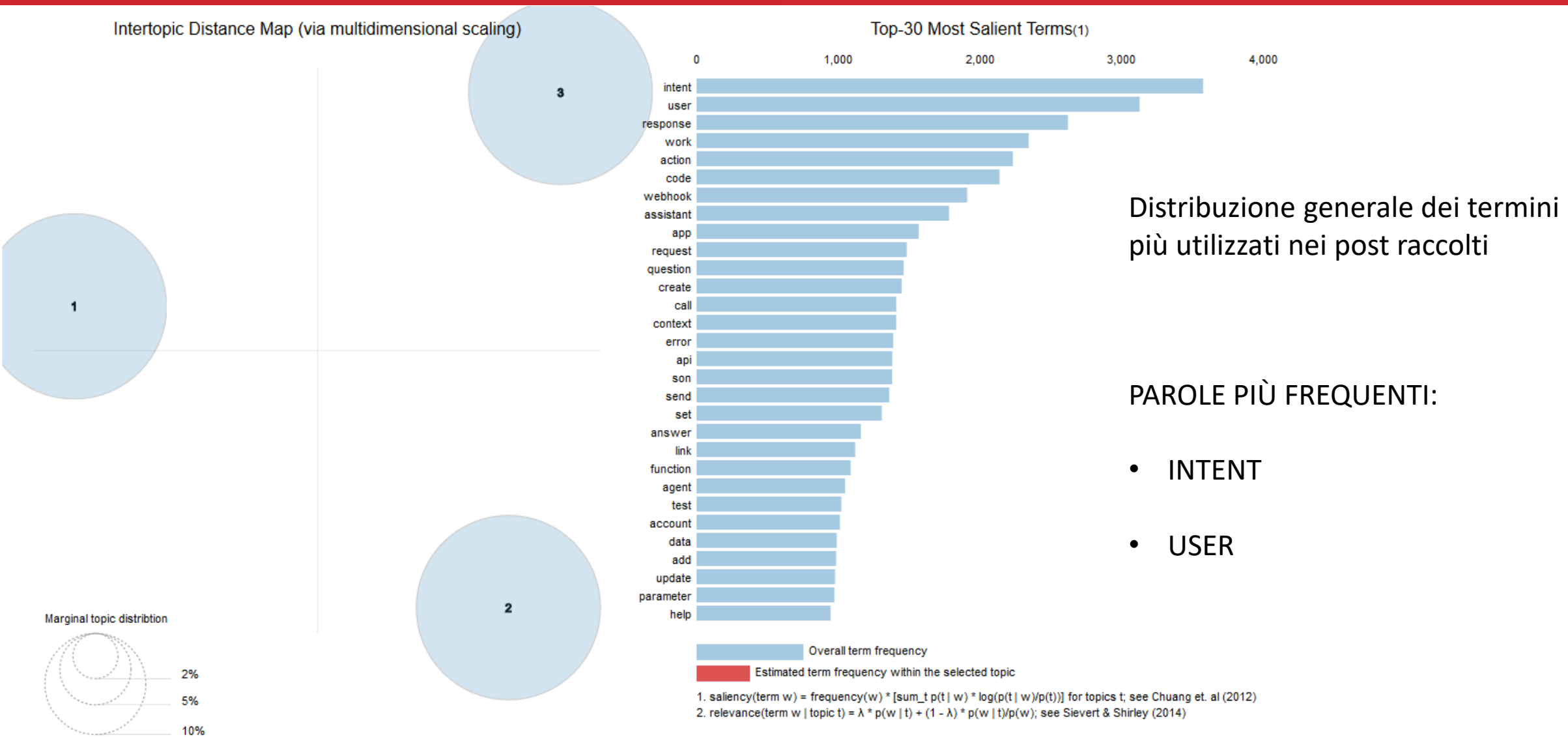


- Nato nel 2008
- Iniziato come una raccolta di vari script Python per la libreria di matematica digitale ceca dml.cz in cui è servito a generare una breve lista degli articoli più simili a un determinato articolo da cui il nome (gensim = "generare simili")
- Oggi software per realizzare la modellazione semantica.

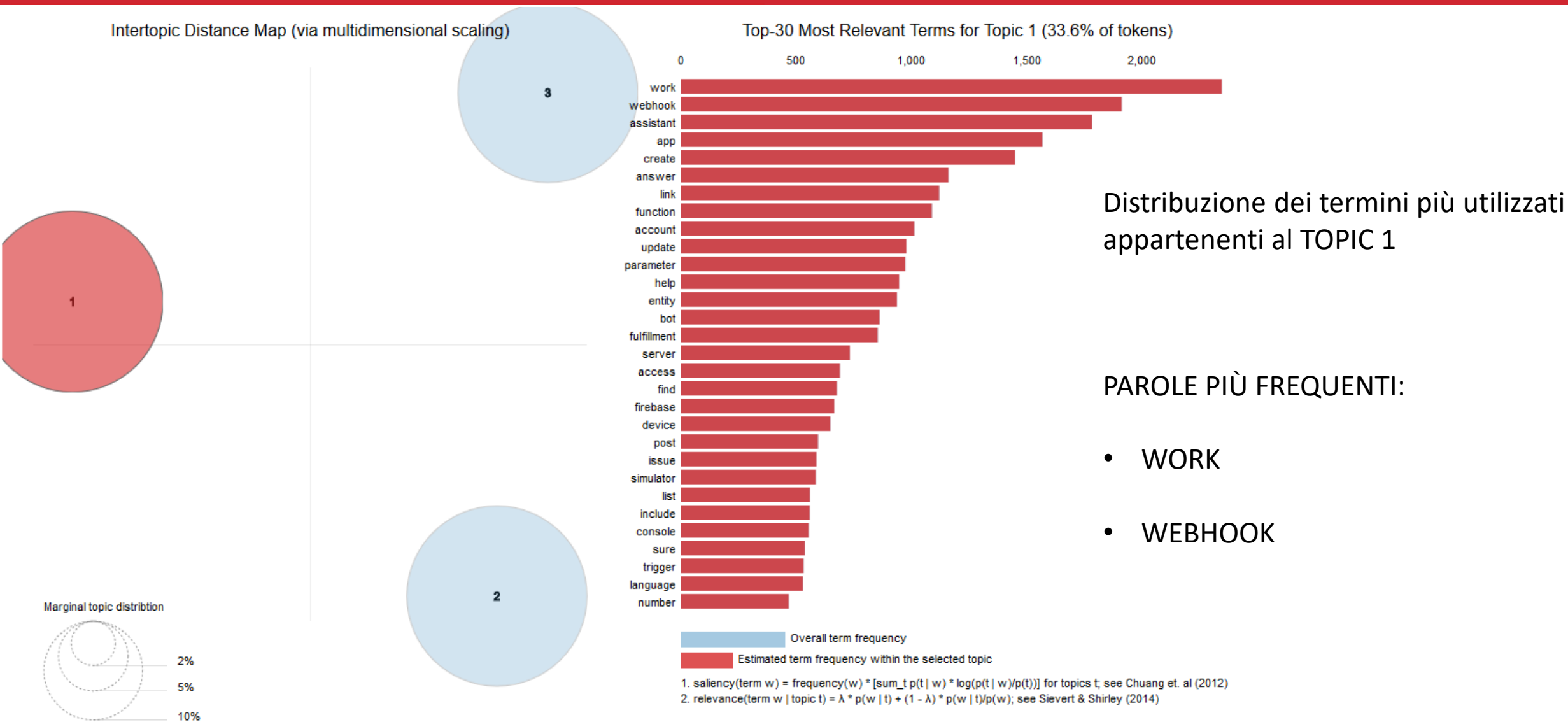
- *Latent Dirichlet Allocation* è un modello probabilistico generativo di un corpus
- L'idea di base è che i documenti sono rappresentati come miscele casuali su argomenti latenti, in cui ogni argomento è caratterizzato da una distribuzione sulle parole



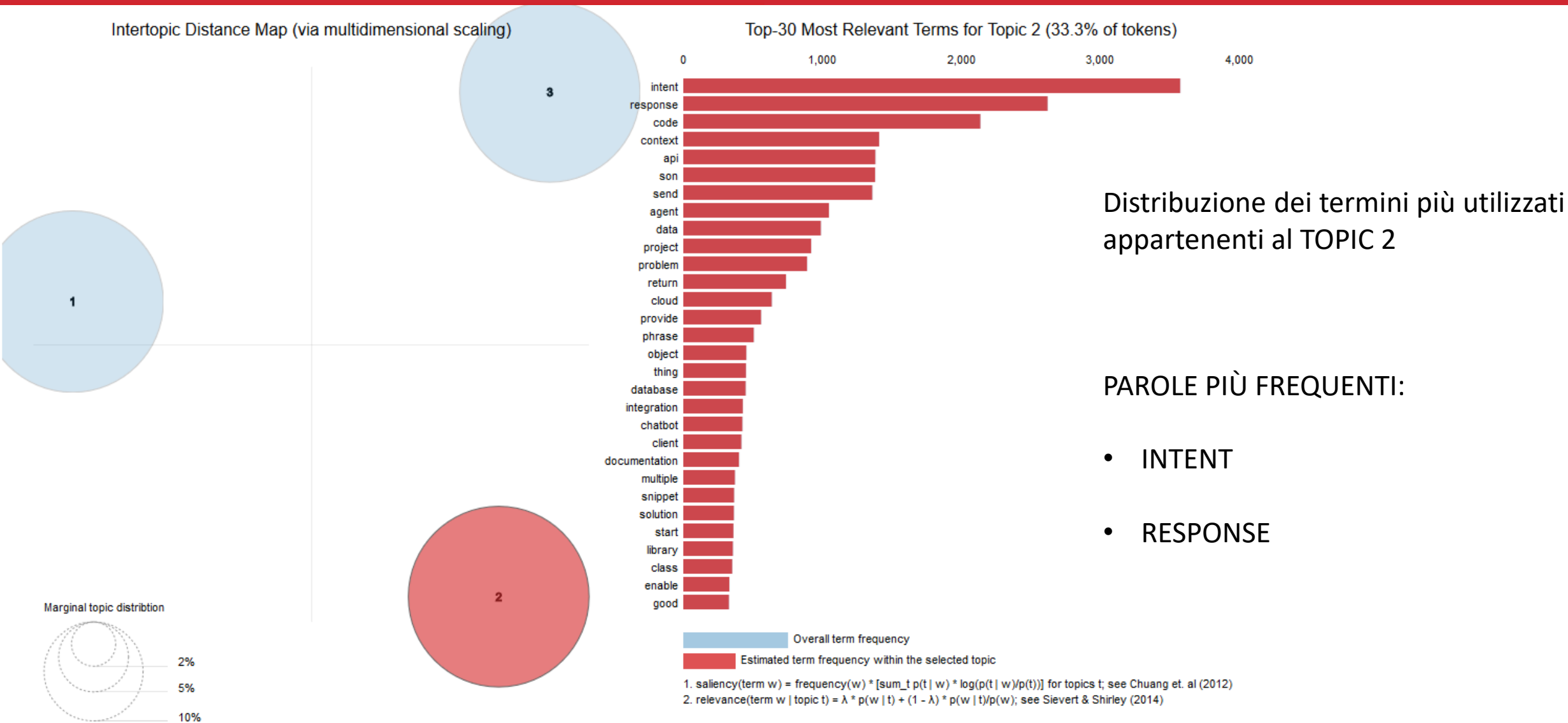
# GRAFICO RISULTANTE DALLA MODELLAZIONE TRAMITE LDA



# GRAFICO RISULTANTE DALLA MODELLAZIONE TRAMITE LDA – TOPIC 1

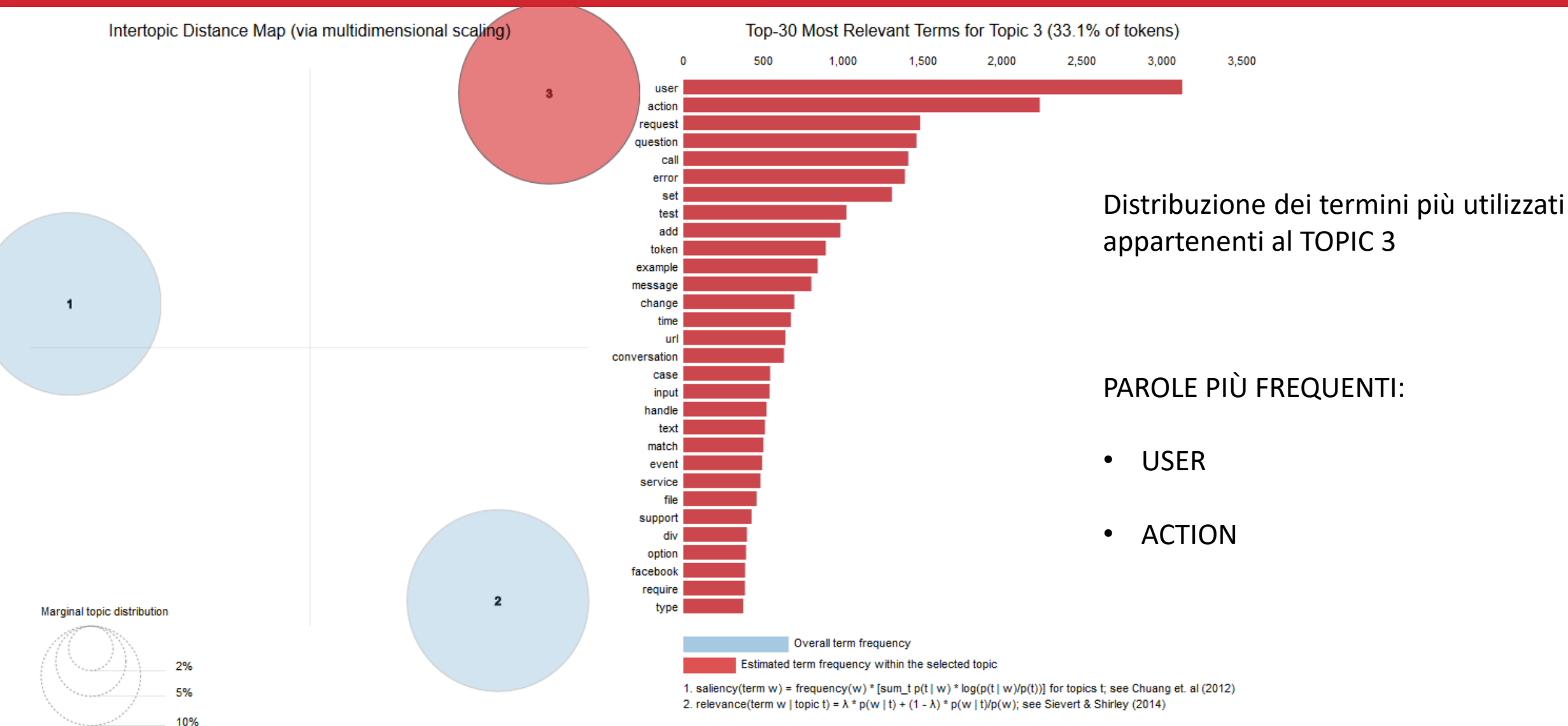


# GRAFICO RISULTANTE DALLA MODELLAZIONE TRAMITE LDA – TOPIC 2

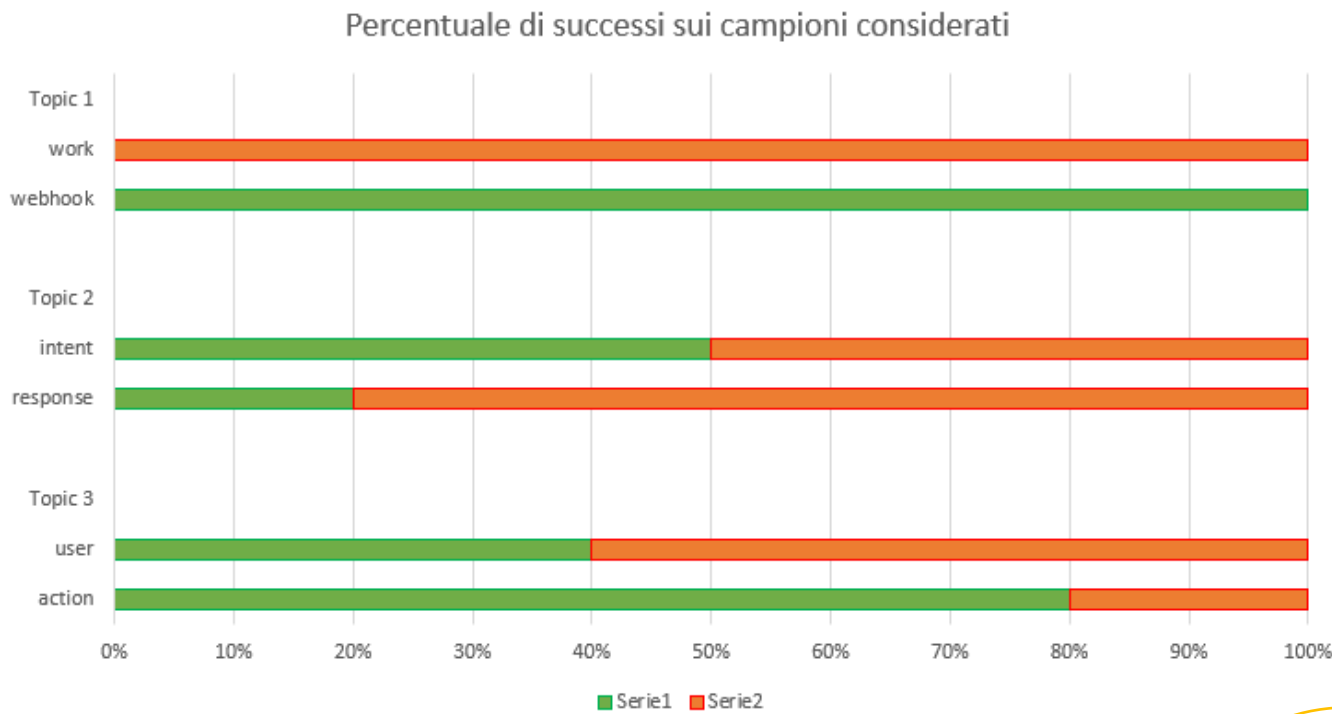




# GRAFICO RISULTANTE DALLA MODELLAZIONE TRAMITE LDA – TOPIC 3



# GRAFICO DEI RISULTATI DELL' ANALISI MANUALE SU THREAD CAMPIONATI CASUALMENTE

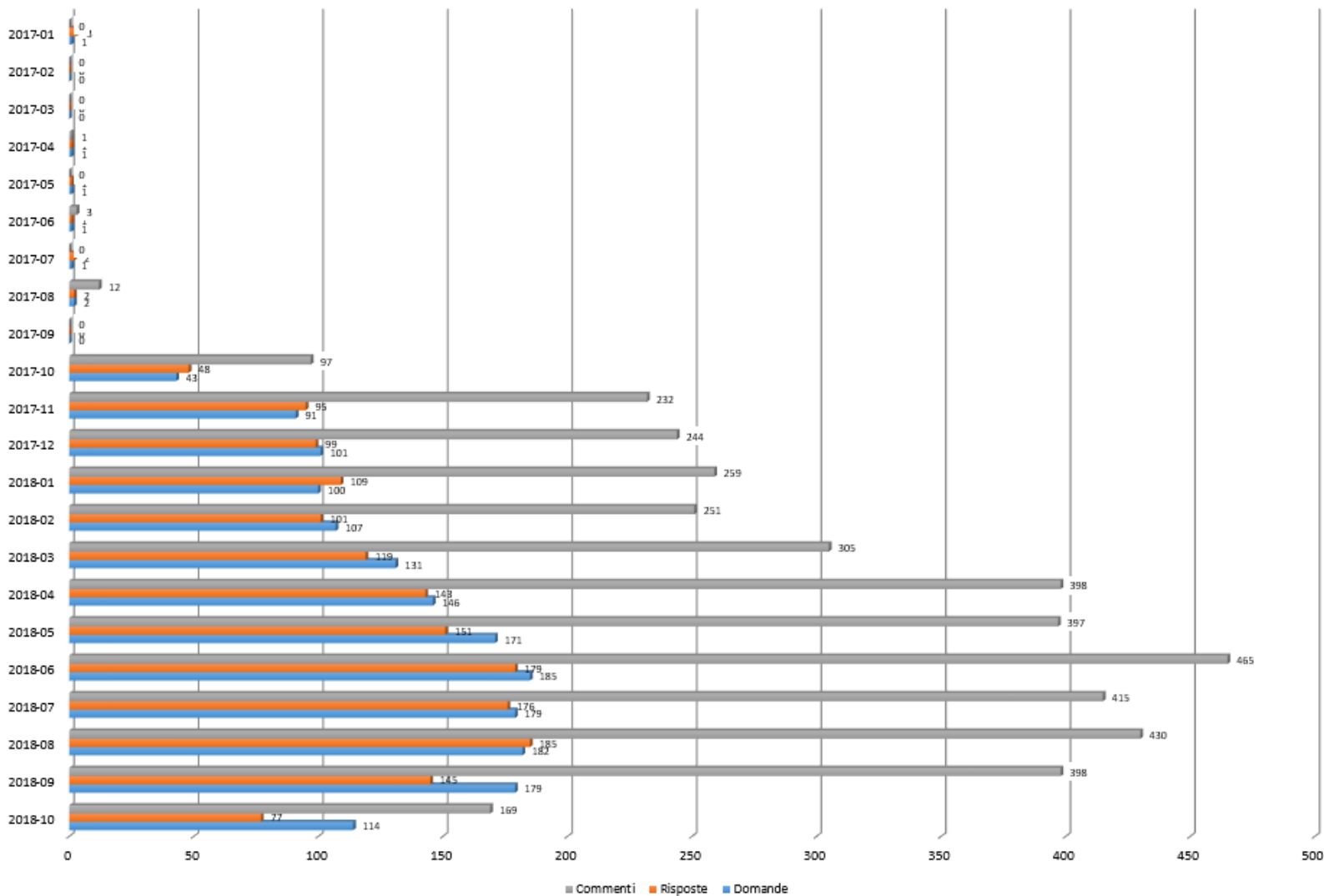


"Now both the codes are not working....."

"The App Actions are the only way to directly do logic between an Android app and an action."

"Intent not providing text ....."

Analisi dei topic di StackOverflow



Ottobre 2017

Dati raccolti

File .csv

# CONCLUSIONI

Durante questo progetto di ricerca ho potuto utilizzare e apprendere Python come linguaggio di programmazione, oltre a migliorare le conoscenze acquisite durante tutto il corso della triennale



StackExchange

Natural Language  
Tool Kit (NLTK)

re — Regular expression operations



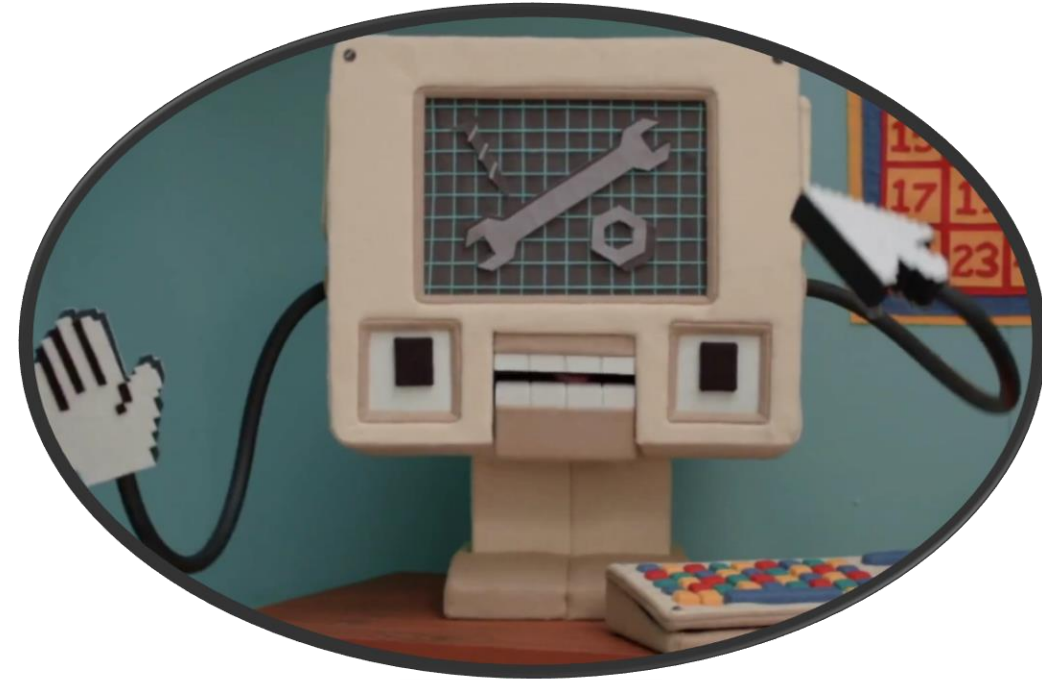
Ho fatto uso anche di molte API e librerie esterne che ho dovuto far interagire e coordinare tra di loro

Come IDE ho utilizzato pyCharm fornito da jet brains



Fondamentale è stato anche l'uso di GitHub per la gestione di backup e versioni del progetto durante tutto l'arco dell'implementazione.

- Ho effettuato una verifica dei dati leggendo casualmente 10 thread in cui compaiono le parole più discusse di ogni topic per verificare con certezza se le problematiche fossero effettivamente legate a questi termini
- Uno sviluppo futuro potrebbe quindi essere quello di automatizzare questa procedura tramite un'analisi, delle sole domande, basata sui dati e i risultati raccolti da questo progetto in modo da estendere questa verifica su l'intero file di dati
- Altrimenti questa operazione sarebbe impraticabile da svolgere manualmente vista la mole d'informazioni



GRAZIE PER L'ATTENZIONE

