



MODULO 4 – INTERPRETARE IN CONTESTO DIGITALE



4.1 IL TEXT MINING: STRUMENTI PER L'ANALISI DI TESTI DIGITALI

4.1 IL TEXT MINING O COME BERE ACQUA DA UN IDRANTE

Perché *mining* (text o data):

- più dati/testi di quanti una persona possa analizzare o leggere in un ragionevole lasso di tempo
- dati/testi che richiedono interventi computazionali perché si possa recuperare un senso

non necessariamente macro vs micro

Usare tecniche, set di dati o metodologie nuove non significa necessariamente scegliere di 'fare in grande'

Obiettivo del text mining: estrazione di informazione in un insieme di documenti e la visualizzazione di grossi set di testi.



4.1. MINATORI IN RETE

**Analogia tra estrazione mineraria e
reperimento informazioni in Internet**

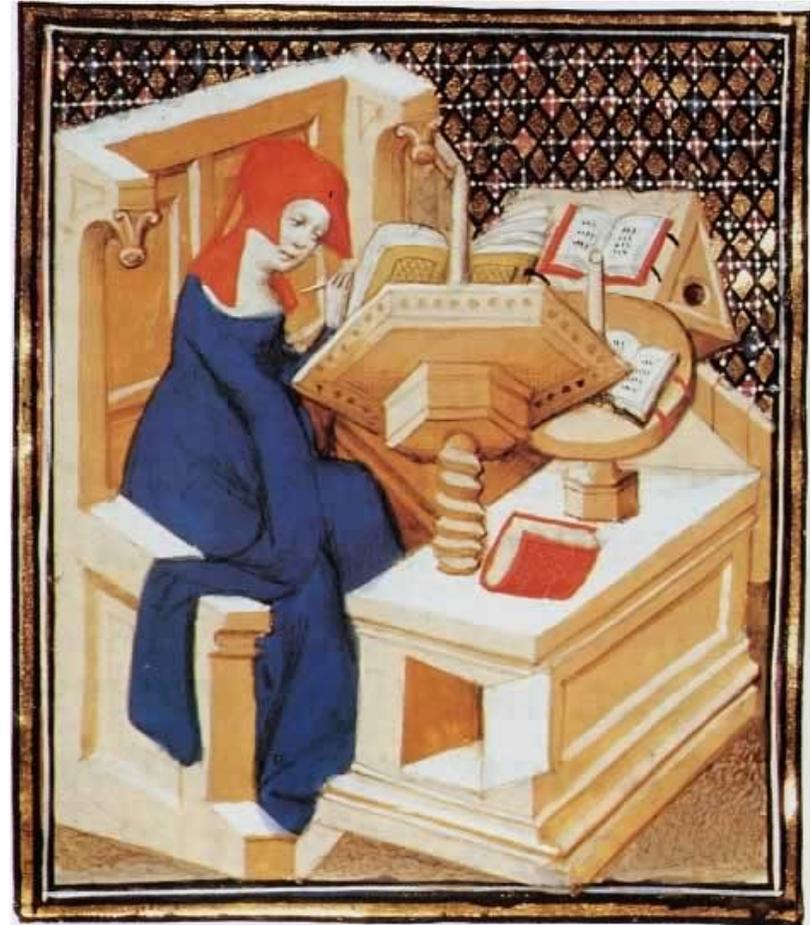
Così come è necessario rimuovere un'enorme quantità di roccia prima che i diamanti o l'oro possano essere trovati, allo stesso modo con il computer e gli strumenti di data mining, possiamo trovare l'unica informazione-diamante tra le tonnellate di dati-roccia nel database.



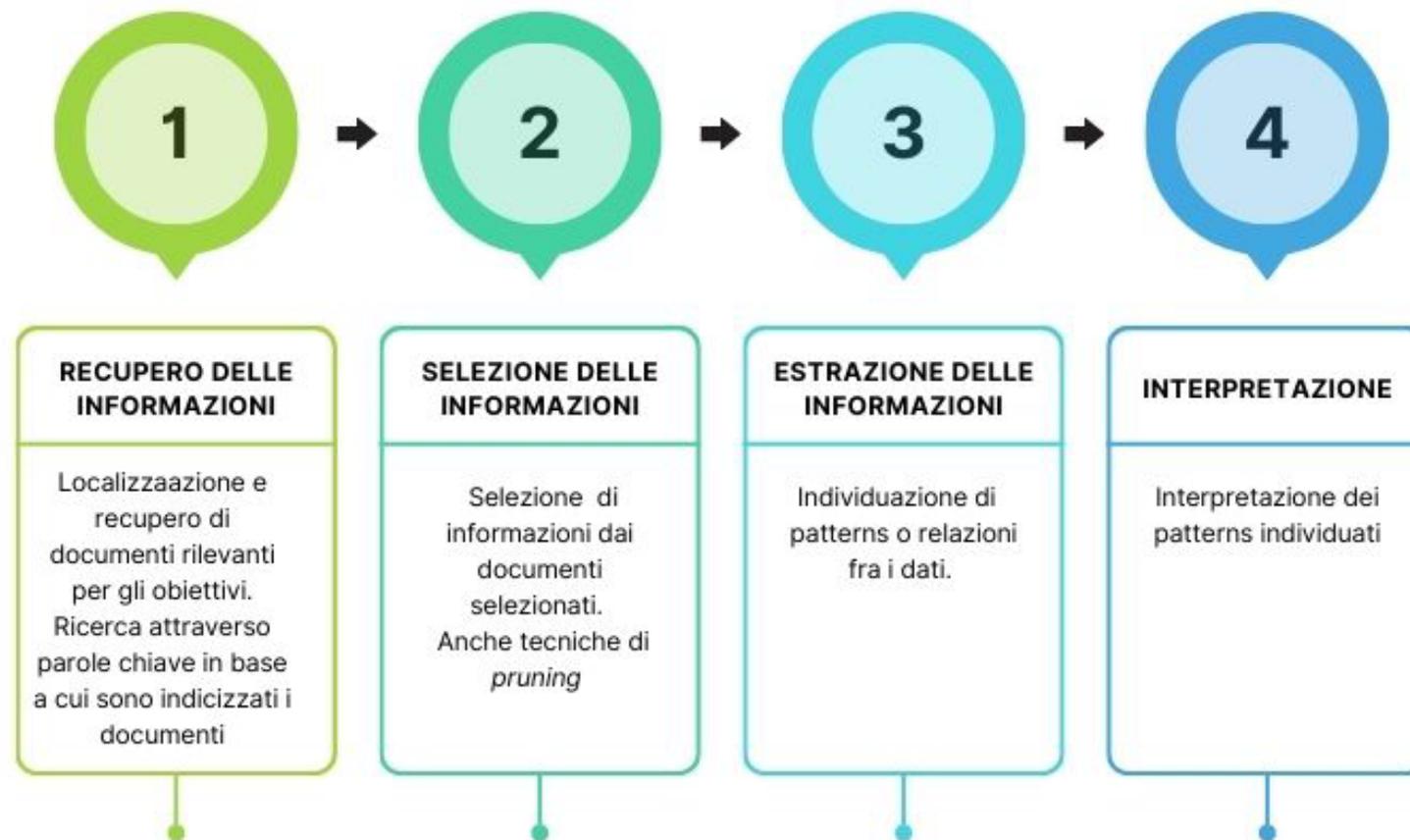
4.1 L'USO DEL COMPUTER NEL TEXT MINING

“As long as there have been books there have been more books than you could read. ... **Knowing how to "not-read" is just as important as knowing how to read**” (Mueller, 2007).

“not-reading” or “distant reading” are the automated **search for patterns** across a much wider corpus than could be read and assimilated via traditional humanistic methods of “close reading.” (Kirschenbaum, 2007)



4.1. IL PROCESSO DEL TEXT MINING



4.1. APPLICAZIONI DEL TEXT MINING

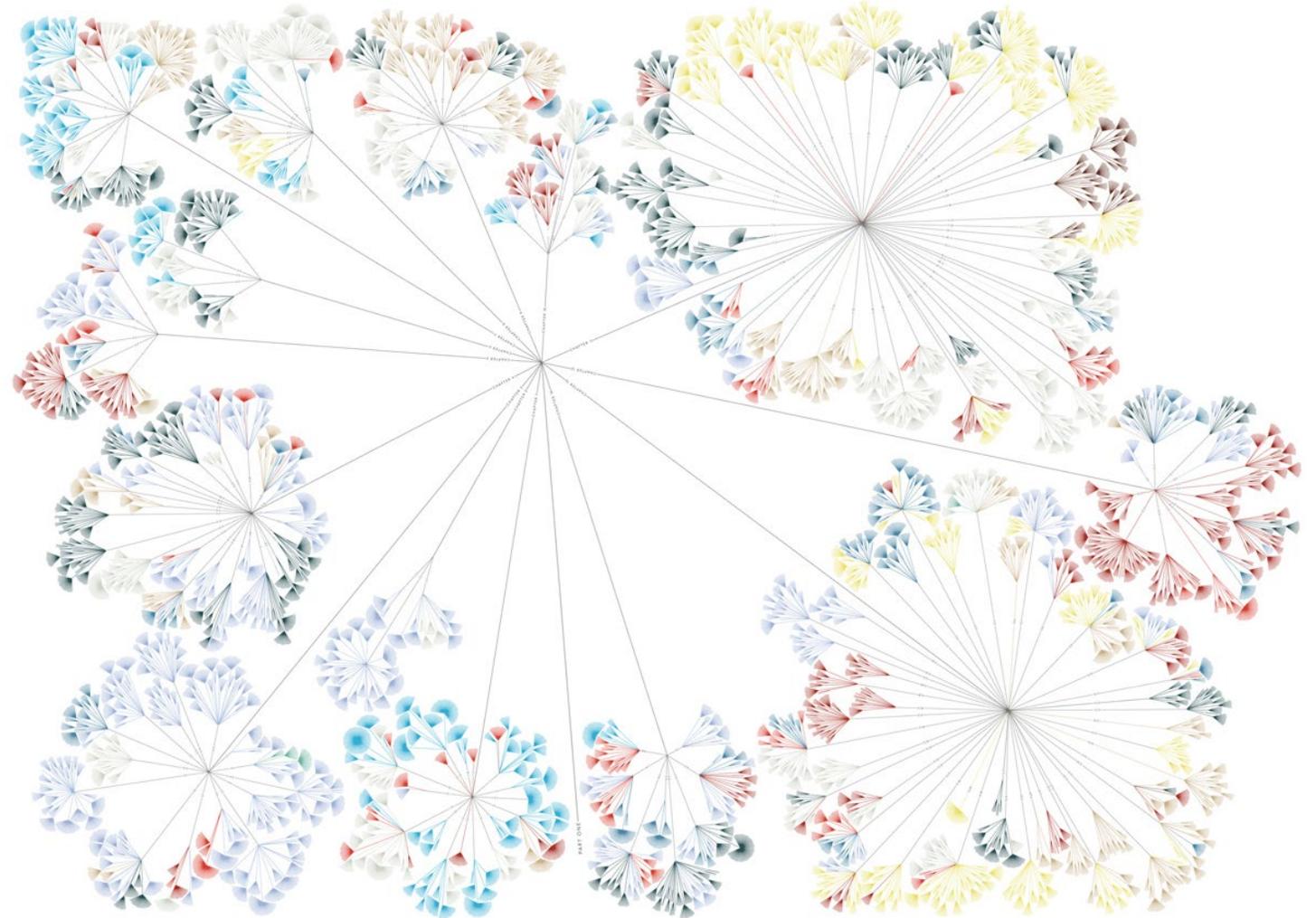
Le tecniche di text mining sono applicabili a qualsiasi ambito di indagine.

In generale trovano applicazione tutte le volte che siamo di fronte a grandi quantità di dati e abbiamo l'esigenza di conoscerne il contenuto.

4.1. DISTANT READING

*Literary Organism –
Posavec*

Distant reading del
romanzo di Kerouac
On the Road



4.1. LA DISTANT READING



I took a course in speed reading and was able to read War and Peace in twenty minutes.

It's about Russia. (Woody Allen)

Possiamo fare di meglio?

4.1. UN ESEMPIO DI DISTANT READING

Google

la locandiera



Cerca con Google

Mi sento fortunato

Circa 1.170.000 risultati (0,41 secondi)

Risultati per 24121 Bergamo BG · Scegli l'area



wikipedia.org

https://it.wikipedia.org/wiki/La_locandiera

La locandiera - Wikipedia

La locandiera è lo standard del nuovo teatro di Goldoni, che soppianta gli schemi logori dell'obsoleta commedia dell'arte. ... Le maschere che gli attori usavano ...

Prima assoluta: 1753

Lingua originale: Italiano

Trama · Lettura · La commedia nel suo contesto... · Note

Le persone hanno chiesto anche

Cosa racconta la Locandiera?



Quali sono i temi principali della Locandiera?



Output in forma di metadati



La locandiera

Commedia di Carlo Goldoni



Anteprima libro

Libro completo disponibile



Leggi ora



Percentuale di utenti a cui è piaciuto questo libro: 64%



Utenti Google

4.1 TECNICHE DI DISTANT READING

Le nuvole di parole > visualizzazioni intuitive per codificare

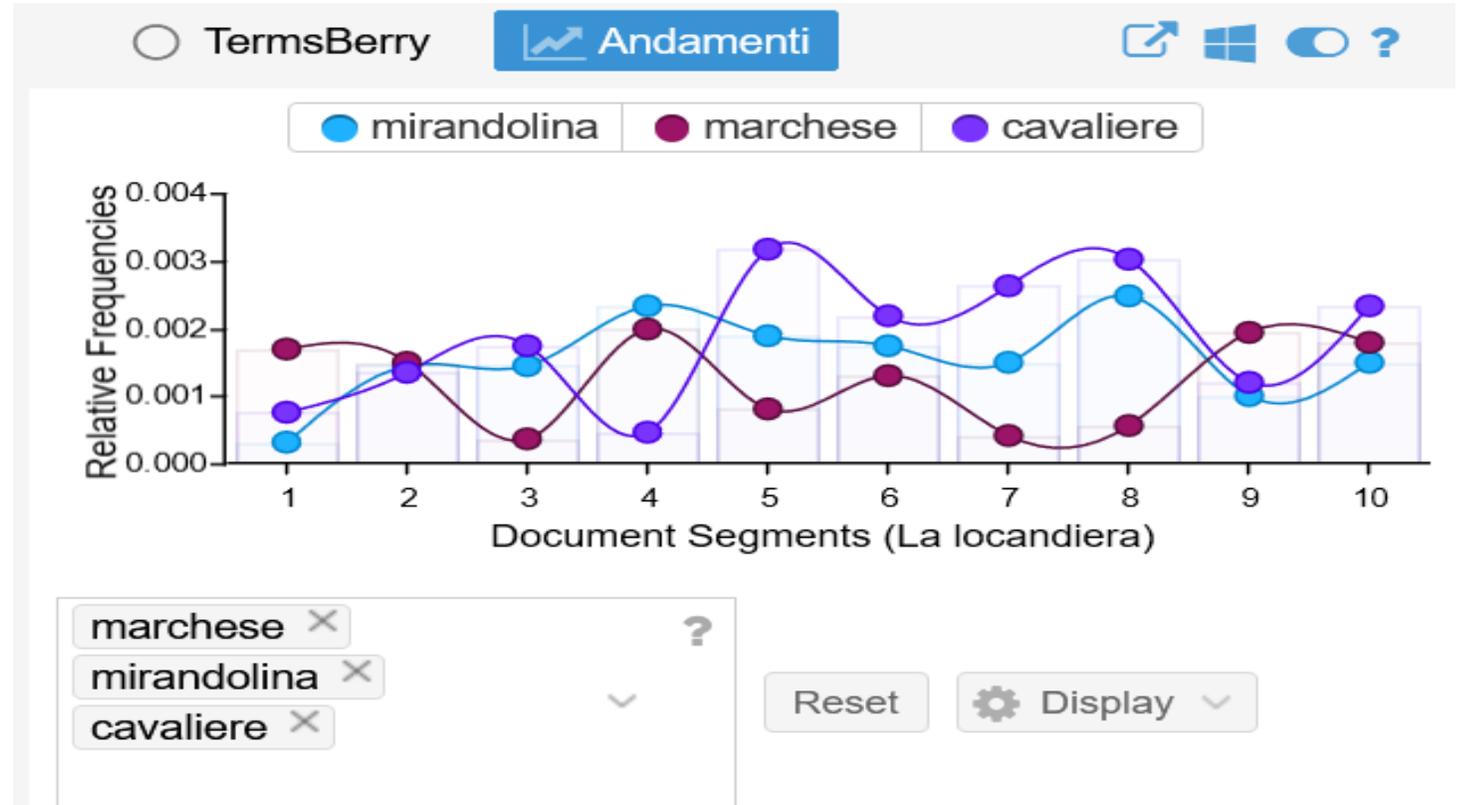
(a) la frequenza di parole in un determinato testo

(b) temi in un corpus

LA LETTURA VELOCE: FREQUENZA PAROLE

	#	Parola	Conteggio	Relativa	Tendenza
<input type="checkbox"/>	1	cavaliere	379	18,708	
<input type="checkbox"/>	1	mirandolina	314	15,499	
<input type="checkbox"/>	1	marchese	247	12,192	
<input type="checkbox"/>	1	conte	190	9,379	
<input type="checkbox"/>	1	fabrizio	115	5,676	
<input type="checkbox"/>	1	signor	93	4,591	
<input type="checkbox"/>	1	signore	63	3,110	
<input type="checkbox"/>	1	signora	27	1,333	
<input type="checkbox"/>	1	illustrissimo	16	790	
			4,298		

POSSIAMO
RECUPERARE
ULTERIORI
INFORMAZIONI
SU 3 PAROLE?



TROVARE TESTI ,SIMILI': CLUSTERING

(ESEMPIO DA GOOGLE NEWS)

- Notizie principali
- Per te
- Stai seguendo
- News Showcase
- Ricerche salvate

- COVID-19

- Italia
- Dal mondo
- Notizie locali
- Affari
- Scienza e tecnologia

Tutta la copertura

la VOCE del TRENTINO
Huggy Wuggy, la polizia postale lancia l'allarme per il pupazzo blu: "E' pericoloso per i bambini"
1 ora fa



il Fatto Quotidiano
Huggy Wuggy, la polizia postale lancia l'allarme per il pupazzo blu: "Contenuto pericoloso per i..."
2 giorni fa



Meteo
Huggy Wuggy: la polizia lancia un avviso, 'Quel pupazzo può creare problemi'; il video
18 ore fa



sky tg24
L'alert della Polizia Postale sul pupazzo Huggy Wuggy: "È pericoloso per i bambini"
2 giorni fa



il Giornale.it
"Non fatelo vedere ai bimbi". Scatta l'allarme sul pupazzo pubblicizzato dagli YouTuber
3 giorni fa



Capoluogo
Huggy Wuggy, il pupazzo blu diventato virale pericoloso per i bambini
2 giorni fa



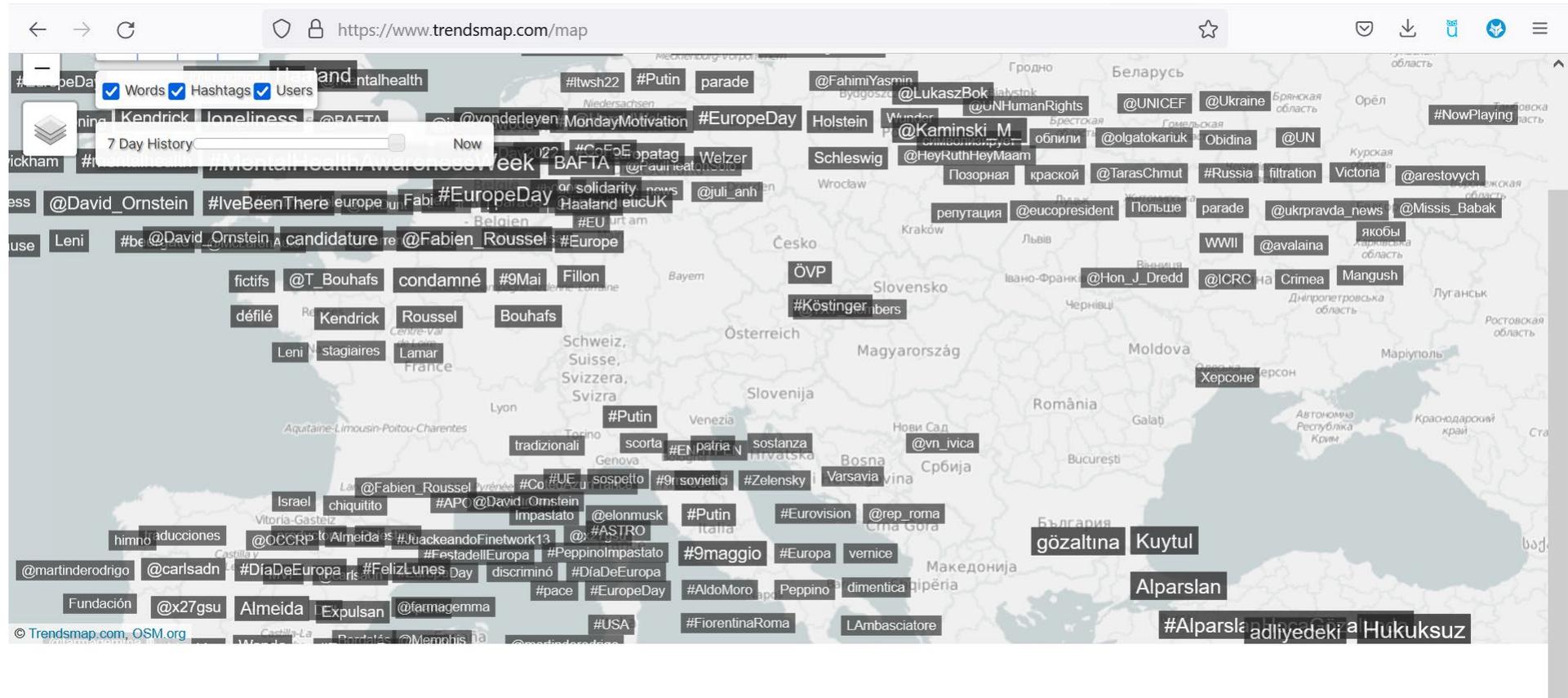
DOMANDE

- Quanto di quello fatto è davvero “automatico “?
- Quali conoscenze (nella mia testa e in quella altrui) sono entrate in questa analisi e come?

4.1 STRUMENTI PER IL DISTANT READING

- 1) Word clouds
- 2) Mappe, soprattutto per informazioni geospaziali contenute in un testo

TRENDING TOPICS: UNA FORMA DI RIASSUNTO



ESEMPI PIÙ INTERESSANTI / SERI

Esempi di Voyant nella ricerca:

<https://voyant-tools.org/docs/#!/guide/gallery>

E alcuni suggerimenti su “da dove iniziare”: <https://voyant-tools.org/docs/#!/guide/start>