

# Post verità, talk della Pasi e del suo assistente

---

## assistente

Sono le notizie diffuse online che fanno appello alle emotività delle persone. Spesso non sono fatti verificati. Si sono diffuse soprattutto durante le elezioni di Trump. Nei social media si crede a tutto, anche a cose non verificate. Infatti i social media raggruppano persone con caratteristiche simili, instaurando tra loro contatti. Si formano reti sociali, *social network*, che sono reti complesse di interazioni, rappresentate da dei grafi. Si ha il **mondo piccolo** (il cammino tra due nodi è al più di 6 archi, *6 gradi di separazione*), con **legami forti/deboli**, il legame forte si ha tra due nodi collegati da un solo arco (possono essere amici e familiari) mentre quelli deboli si basano su interessi in comune etc... Il **numero di Dumban**, 150, indica il numero massimo di interazioni forti possibili. Alla fine gli utenti si trovano nelle **camere d'eco**, dove tutti la pensano ugualmente e dove una certa teoria risulta inconfutabile. Alcuni algoritmi sui social spingono verso le camere d'eco, sono gli **algoritmi di filtraggio**, ci si trova quindi in una **bolla di filtro**. Una tecnica in uso è il **fact checking**, per verificare le notizie... ma anche questo viene rifiutato dalle camere d'eco. Si cerca quindi un modo per verificare la credibilità delle informazioni, i cui studi sono iniziati negli anni 50, valutando affidabilità e competenza. I social media hanno generato disintermediazione, quindi tutti possono postare contenuti senza moderazione o controllo. La fonte dell'informazione non è più verificabile. Si hanno recensioni false (*opinion spam*), fake news e la credibilità dell'informazione medica. L'**opinion spam** può comportare il mutare l'opinione di altri utenti su un certo servizio. Se si ha una recensione negativa spesso non si effettua un acquisto. Lato **fake news** si ha l'indirizzamento dell'opinione pubblica; appaiono spesso nei **micro-blog**, come **Twitter**, dove si hanno brevi testi che si diffondono velocemente, sono i **trending topics**. Per la credibilità dell'informazione medica si ha che molti cercano medicine e cure online, con risultati disastrosi. Si vuole aumentare l'alfabetizzazione sanitaria. Vediamo un po' le tecniche per verificare la credibilità. Si parte dai dati, **data science**, analizzandoli per trovare "indizi". Un'altra branca di approcci si basa sulla **conoscenza pregressa di dominio**, che si basa, per esempio, sull'uso di linguaggio tecnico etc... Ci sono approcci che verificano come si propaga una notizia vera o falsa, bloccando gruppi che propagano nozioni false, oppure che classificano le notizie sulla base del **livello di verità**, assegnando una **percentuale di verità**. Si usa anche il **machine learning**, un'ingente mole di dati viene campionata in sottoinsiemi di cui si sa la veridicità e lo si usa per estrarre la veridicità, mediante la costruzione di un modello, dagli altri sottoinsiemi. Ogni notizia ha delle caratteristiche di credibilità, data dall'utente che la scrive e dal suo profilo sui social. Si sta lavorando sulla limitazione delle *camere d'eco*, su come fornire informazioni all'utente con nuovi metodi e sul riconoscere e bloccare subito eventuali *fake news*, come per esempio

con **askPinocchio**, un *detector* di fake news molto basilare basato sulla rete sociale di una certa notizia.

## Pasi, laboratorio di Information Retrieval

Si usa il **text mining**, estraendo informazione dai testi. **Internet** è una rete di reti che collega vari device fisici mentre il **World Wide Web** è una sovrastruttura di internet che permette la condivisione di contenuti in maniera più semplice, prima si passava per *ftp* per esempio, mediante la creazione di pagine web, si ha quindi la creazione dei **browser**. Si utilizzano ipertesti con *link* per collegarsi alle pagine. Poi si è avuto il **web 2.0**, con i social media, dove chiunque può dire la propria in maniera istantanea. Si hanno sistemi per la **raccomandazione di informazioni**, visiti un contenuto e poi le pubblicità ti bombardano con quel contenuto. I suggerimenti arrivano per esempio tramite i **cookie** e le infinite tracce digitali che lasciamo online. I **motori di ricerca** nascono nel '94, **google** nel '98 e all'inizio erano prodotti venduti commercialmente (i non web dagli anni 70). I motori di ricerca sono diversi da un sistema di **basi di dati**, che estraggono un dato esatto con una *query*, in quanto si fa un'analisi del testo dato in input con dei risultati plausibili, in parte diversi da persona a persona, che tengono conto anche del contesto di chi fa la query (negli ultimi anni). Google è un **motore di ricerca verticale**, che permettono la ricerca su tipi di contenuti, pagine, audio etc... I motori di ricerca sono la punta dell'iceberg dell'**Information Retrieval**. La query arriva ai server di Google dove vengono confrontate le **keyword**. Ogni motore ha un suo algoritmo di **crawling** proprietario, poi ad ogni pagina si applica un **algoritmo di indicizzazione**, che estrae le singole parole dalle pagine, che verranno confrontate con la query dell'utente. Il web è rappresentabile con un grafo orientato. Si ha l'algoritmo di **page rank**, che stabilisce le pagine più gradite dagli utenti e viene usato per ordinare i risultati delle ricerche, insieme ad altri fattori come la localizzazione geografica e l'uso di machine learning.

## Tracce Digitali, Pasi

La **ricerca personalizzata** può essere comoda per la qualità dei risultati ma si ha anche un aspetto negativo dato dalla vendita delle informazioni (ma per esempio **Google** non può farlo, non vende ma analizza, a differenza di Facebook, per esempio, che ha venduto e continuerà a farlo). Google, una volta effettuato l'accesso, sfrutta la cronologia delle ricerche e l'uso degli altri applicativi Google per **personalizzare i risultati**. Ogni nostra azione sul web lascia tracce digitali, migliaia di migliaia di tracce digitali. Anche il sistema a **like, review** e **follower** dei social media è fortemente usato in questo sistema di tracciamento al fine di profilare l'utente. Bisogna aggiungere che la ricerca personalizzata spesso limitano la scoperta casuale. Si sta sviluppando anche un sistema per la **medicina personalizzata**, con un grande occhio per la privacy e la qualità dei risultati.