

Relazione ModSem - FaktNews

Pio Raffaele Fina, Lorenzo Tabasso
pio.fina@edu.unito.it,
lorenzo.tabasso@edu.unito.it

Dipartimento di Informatica, Università degli Studi di Torino

Sommario La studio della verità è uno dei principali argomenti di interesse della filosofia, costituendo da millenni, tema di dibattito filosofico. Nella *teoria neo-corrispondentista*, il concetto di verità si basa essenzialmente su un argomentazione ontologica: una credenza¹ è vera se esiste un'entità appropriata - un fatto nel mondo - a cui corrisponde [1]. Di conseguenza, se non esiste tale entità nella realtà, allora tale credenza può essere ritenuta ragionevolmente falsa.

Il fenomeno globale delle *Fake News*, e più in generale quello della disinformazione, cresce e si diffonde a ritmi esponenziali, mettendo a rischio la stabilità di molteplici aspetti della società in cui viviamo. Una possibile interpretazione è data nel riconoscere in una fake news uno strumento per la generazione di una **realità alternativa**; una pseudo-realità con un certo grado di verosimiglianza, ma falso nella corrispondenza dei fatti. Partendo da questo presupposto, non risulta difficile immaginare, come "lo strumento fake news" possa risultare una concreta e attuale minaccia da molteplici punti vista, in primis, la credibilità del giornalismo e la stabilità democratica di uno Stato.

Sebbene il termine fake news sia diventato *mainstream* solo di recente, parola dell'anno 2017 [2, 3, 4], la diffusione intenzionale di informazioni false o tendenziose, è un fenomeno da ricercare sin dall'avvento dei primi sistemi di scrittura e di stampa [5, 6]. Esempi di rilevo e ad alto impatto sociale sono state le attività di propaganda durante la I e II Guerra Mondiale, la Guerra Fredda e le guerre jugoslave.

Il fenomeno delle fake news qui analizzato, assume connotati differenti, e per questo motivo deve essere ri-contestualizzato nell'attuale scenario comunicativo. La digitalizzazione delle notizie e la diffusione di piattaforme di *social media* e *instant messaging* ha radicalmente trasformato il panorama informativo e la definizione tradizionale di notizia: un *tweet*, un breve messaggio di 140 caratteri, viene considerato a tutti gli effetti un "pezzo d'informazione", in particolare se proviene da una fonte autorevole [5]. Nata in origine come piattaforme in cui gli utenti potevano condividere le proprie opinioni personali e rimanere in contatto con amici, si sono trasformate in portali dove gli utenti producono, condividono e consumano informazione, spesso sotto forma di notizie. La facilità, la rapidità di condivisione e la vastità dell'audience a disposizione facilitano lo scambio e diffusione di notizie, ma allo stesso modo anche di fake news[5][7, ch. 3]. Per capire il ruolo dei social network nella diffusione delle fake news si può provare un'interessante applicazione interattiva: getbadnews.com.

Il termine, ora ampiamente abusato, descrive storie interamente fabbricate e dal contenuto sensazionalista che raggiungono una vasta audience sfruttando i social media a proprio vantaggio [7, ch. 4].

Data l'elevata rilevanza del problema, molteplici iniziative di *fact-checking* (transnazionali e non) sono nate nel corso dell'ultimo decennio per contrastare questo fenomeno, esempio di notevole rilievo e autorevolezza è l'*International Fact-Checking Network*. A differenza dell'accezione del termine riferito al passato, l'attività attuale di *fact checking* viene svolta *ex-post* alla pubblicazione di una notizia o dichiarazione. L'obiettivo principale risiede nel **verificare** la veridicità dell'affermazione ricercando **fonti** fattuali che supportano o confutano una data affermazione.

¹ qui da intendere come *belief*

Convenzioni Per facilitare la lettura, nel documento viene utilizzata la seguente formattazione:
`<prefix>:<SomeClass>` , `<prefix>:<someObjectProperty>` , `<prefix>:<someDataProperty>` per indicare rispettivamente una owl:Class, owl:ObjectProperty, owl>DataProperty. Per i diagrammi viene utilizzata la stessa convenzione della documentazione di Provenance. Il testo in `text` indica un riferimento ad aspetti implementativi legati principalmente al codice.

1 Descrizione del Dominio

Il dominio scelto, seppur condivide molti degli aspetti del giornalismo investigativo digitale, presenta alcune peculiarità proprie del fact checking.

Le principali entità del dominio sono i *claimants*: coloro che effettuano una certa dichiarazione su un qualche canale di comunicazione; e i *fact-checkers*: le persone fisiche, organizzazioni e agenti software che controllano la veridicità delle affermazioni fatte. Spesso i fact-checkers sono affiliati ad *organizzazioni* che godono di un certa credibilità e autorevolezza, caratteristiche essenziali per la fiducia da parte dei *lettori*. In alcuni casi, queste organizzazioni corrispondono a redazioni giornalistiche (es. [Open.online](#) e il [Washington Post](#)) che avviano dei progetti paralleli rispetto al piano editoriale principale.

La redazione monitora le principali testate giornalistiche, agenzie stampa e profili social dei personaggi pubblici, selezionando principalmente le dichiarazioni basate su fatti verificabili (e quindi non opinioni). Il risultato del processo investigativo del fact-checker consiste nella realizzazione di un articolo di *review*.

Questi articoli, seppur altamente variabili nei contenuti, presentano alcuni elementi ricorrenti nella forma. Inizialmente nel testo viene sempre descritto il contesto in cui viene effettuata una certa dichiarazione. Inoltre viene sempre riportata la fonte della dichiarazione, come contenuto ipertestuale esterno deferenziabile. Il contenuto principale dell'articolo decompon e analizza la dichiarazione, esponendo le varie evidenze trovate che supportano o confutano un certo punto. L'articolo, nella quasi totalità dei casi, termina con un *giudizio* sommario dell'autore accompagnato da un valore di *rating* che esprime quanto la dichiarazione sia veritiera e quindi supportata dai fatti.

Di fondamentale importanza è il ruolo ricoperto dalle *evidenze*, documenti digitali multimediali sempre deferenziabili e quindi scrutinabili dal lettore. La scelta dell'evidenza è critica non solo dal punto di vista del contenuto ma anche dalla *fonte* da cui esse provengono. Le linee guida delle organizzazioni tendono sempre a scegliere fonti di evidenza affidabili e autorevoli, ad es. NGO internazionali, istituzioni governative e gruppi di esperti.

Una caratteristica del dominio che sta emergendo, e che di conseguenza abbiamo voluto modellare, consiste nel considerare anche gli agenti software basati su recenti tecnologie generative come parte attiva delle dinamiche del dominio; esempi recenti sono [Deep Fakes](#), [Trump Bot](#) e [OpenAI GPT3](#).

2 Requisiti Ontologici

In questa sezione sono riportati i principali requisiti individuati. Il dominio scelto riguarda un'area ristretta del giornalismo investigativo digitale: il fact-checking. Nello specifico abbiamo ristretto il campo d'interesse sul **processo di fact-checking** di dichiarazioni politiche². Le motivazioni alla base della scelta del dominio sono molteplici.

Come già riportato nell'introduzione, la disinformazione è un tema ad alto impatto sociale, soprattutto se si considera il ruolo di *decision maker* dei politici all'interno della società.

L'articolo di review rappresenta l'output del processo di fact-checking, processo che solitamente viene effettuato secondo principi etici e d'imparzialità. A dispetto dell'impegno e della sottoscrizione ad un determinato codice etico, questo non esclude che la valutazione effettuata dal fact-checker, possa presentare possibili bias cognitivi. Due delle colonne portanti alla base del giornalismo sono la **trasparenza** e la **fiducia** [8][9], per questo motivo, crediamo fortemente che l'intero processo dell'attività giornalistica debba essere attribuibile, **tracciabile**, spiegabile e **verificabile** dai lettori, pertanto la nostra modellazione non si limita esclusivamente all'entità "review" ma include anche la strutturazione dell'intero processo di fact checking.

² Sebbene il campo d'interesse sia stato ristretto ad una particolare area, la struttura dell'ontologia proposta, non vincola in nessun modo l'utilità in altre aree.

Inoltre, la maggior parte dei progetti legati al fact-checking con tecnologie basate sul *Semantic Web*, hanno principalmente come oggetto d'interesse la review [10][11] o la realizzazione di tecniche di *fake news detection*; ponendo poca o nessuna enfasi sulle dinamiche del processo. Pertanto il nostro lavoro nasce come proposta per colmare questo gap nell'attuale panorama.

Task, Utenti e Scenari I principali utenti individuati sono:

- **Fact checker:** un qualsiasi agente software, persona fisica, o organizzazione che effettua una delle attività di fact checking. Nella maggior parte dei casi corrisponde all'autore di un articolo di fact checking.
- **Lettore:** persona fisica che legge un articolo di fact checking, mossa dall'interesse di verificare la veridicità di una qualche dichiarazione.
- **Editor:** persona fisica il cui ruolo principale è assegnare i contenuti ai fact-checkers. Spesso verificano la validità e approvano la pubblicazione dell'articolo di review scritto dai fact-checkers.

Nell'analizzare il dominio sono stati individuati due principali scenari d'utilizzo:

- **Auditing e Verifica:** l'ontologia permette di tracciare le principali attività svolte da differenti individui con il relativo ruolo ricoperto.
In questo scenario, ipotizziamo che gli utenti maggiormente interessati siano il *lettore* e l'*editor*. Il primo può accedere ad una serie di informazioni a monte del processo che solitamente rimangono "nascoste" ed implicite. Sebbene queste informazioni siano di principale interesse degli "addetti ai lavori", esse ricoprono un ruolo fondamentale nel garantire trasparenza e fiducia.
Grazie alle informazioni codificate, l'*editor* può verificare che l'operato svolto dai suoi collaboratori sia conforme alle *policy* e linee guida dell'organizzazione.
- **Annotazione Semantica** l'ontologia (in alcune delle sue parti) permette di assegnare alle entità d'interesse in una review la loro descrizione semantica utile per gli strumenti di information retrieval.
In questo scenario, ipotizziamo che il principale utente sia il *fact-checker*. Durante l'intero processo e in particolare durante la stesura della review può annotare manualmente o aiutato da strumenti semi-automatici l'intero articolo, in modo da creare incrementalmente la base di conoscenza dell'organizzazione, realizzando i principi dei *Linked Open Data* (LOD).

L'ontologia proposta può essere utilizzata come "backbone data layer" di differenti sistemi informativi. Molte delle istanze e object properties dell'ontologia potrebbero essere manualmente inserite tramite wizard guidati o utilizzando metodi automatici di information extraction.

3 Documentazione e Modellazione del Dominio

Parafrasando il famoso aforisma di George Box "*tutti i modelli sono errati, ma alcuni sono utili*", di seguito, vengono presentati i principali aspetti, assunzioni e semplificazioni effettuate nel processo di modellazione dell'ontologia:

- **Rating:** data l'inerente ambiguità del linguaggio naturale con cui sono espresse le dichiarazioni, stabilire la veridicità di una notizia, non può essere ridotta ad un semplice valutazione dicotomica "notizia vera" o "notizia falsa". In un contesto reale, risulterebbe in un'assunzione molto forte da effettuare. Di conseguenza, basandoci sui vari schemi utilizzati da altre organizzazioni (Figura 1), utilizziamo una **scala di rating** a 4 livelli che quantifica l'affidabilità e la credibilità di una affermazione.
- **Granularità dell'informazione:** alcune parti di una dichiarazione potrebbero essere vere e supportate da fatti, altre false, altre ambigue e altre parzialmente vere. A differenza dell'approccio utilizzato in [10], non abbiamo voluto modellare l'informazione contenuta in una dichiarazione o Review da un punto vista linguistico. Il nostro modello non permette di catturare e annotare l'informazione espressa in uno specifica posizione in un testo (o video).
- **Bias:** sebbene i principi dettati dall' IFCN costituiscono un chiaro codice etico per l'attività di fact checking (Figura 2), l'attività svolta dai Fact Checkers potrebbe comunque essere soggetta a bias psicologici degli stessi. Non abbiamo voluto modellare questo aspetto in quanto avrebbe richiesto uno studio approfondito di queste dinamiche che avrebbe spostato il focus del progetto. Piuttosto, con l'ontologia proposta, forniamo agli utenti interessati la possibilità di accedere all'informazione necessaria in modo da poter trarre conclusioni personali.
- Il modello proposto non cattura l'idea che una fake news può essere interpretata come manifestazione di molteplici notizie e dichiarazioni, come espresso dal modello concettuale **FRBR**. Anche se questo ricopre un importante punto di vista, il nostro focus di modellazione è il processo di Fact Checking; processo che si realizza su una particolare manifestazione della fake news, piuttosto che la stessa in termini astratti.
- **Stato del Mondo:** seppur banale è importante esplicitare un'assunzione fondamentale alla base del nostro modello: il giudizio di un Fact Checker, si basa sullo " stato del mondo" al tempo di scrittura dell'articolo. Se questo cambia nel futuro, il giudizio è da ritenersi comunque valido in base ai fatti di supporto disponibili al Fact Checker in quel particolare contesto temporale.

Alla luce di queste osservazioni, il modello proposto è basato sulle descrizioni delle metodologie delle principali iniziative di fact checking a livello internazionale [12, 13, 14, 15]. Vogliamo evidenziare che allo stato attuale, non esiste un vero e proprio standard internazionale che regolamenta il processo. Esistono però delle realtà, iniziative e organizzazioni, le cui metodologie condividono la maggioranza dei principi etici e convergono a delle "best practices", spesso basate sul *IFCN code of principles* (Figura 2).

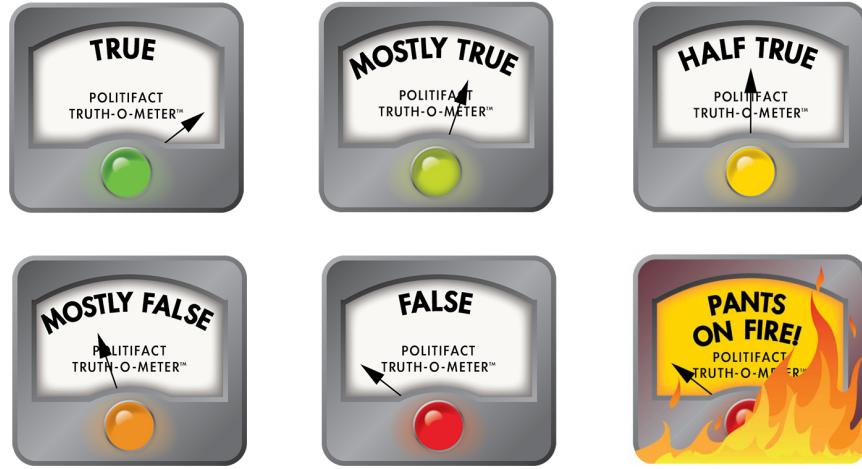


Figura 1: Rappresentazione grafica del "Truth-O-Meter", la scala di rating a sei livelli utilizzata da [Politifact.com](#)



Figura 2: Alcuni dei principi proposti nel *IFCN Code of Principles*. Ulteriori informazioni si possono trovare al seguente [link](#)

Il modello concettuale alla base dell'ontologia (Figura 3) decomponge il processo di fact checking in 3 attività principali:

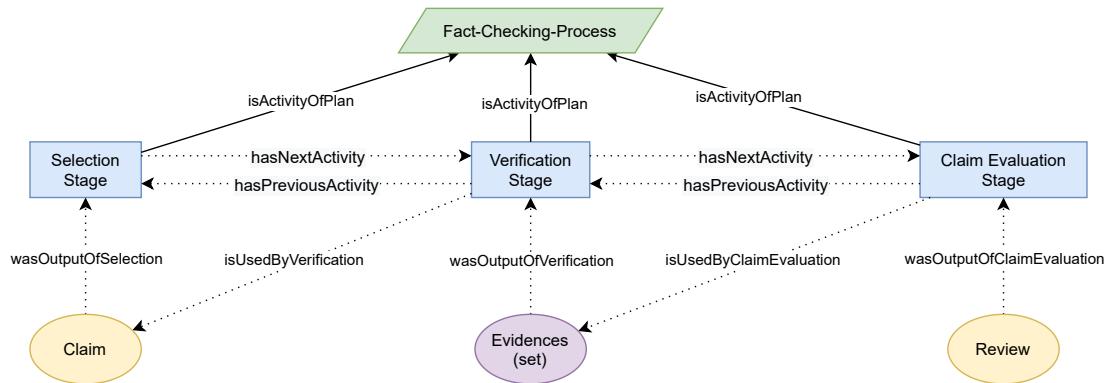


Figura 3: Modello concettuale del processo di fact checking

- fn:SelectionActivity**: in questa fase uno o più **prov:Agent** stabiliscono quali **fn:Claim** selezionare da una possibile **fn:AlternativeList**. La decisione è dettata principalmente dalla linea editoriale definita dall'**fn:Organization** a cui il **fn:FactChecker** è affiliato (**fn:affiliatedTo**). Dal punto di vista della modellazione è interessante notare come modellare esplicitamente la lista di alternative permette di catturare due dinamiche: il lettore può valutare quali decisioni vengono prese a monte dall'organizzazione, in modo tale da analizzare eventuali presenze di imparzialità (si pensi a faziosità politiche); inoltre permette ad altre organizzazioni di verificare quali statement sono stati già revisionati e quali no, creando così una "repository condivisa" e un area di lavoro comune.
- fn:VerificationActivity**: in questa fase il Fact Checker svolge un'attività investigativa di ricerca volta ad individuare le **fn:Evidence** che confermano (**fn:SupportingEvidence**) o confutano (**fn:DisprovingEvidence**) il contenuto del Claim. Di particolare interesse in questa fase è la metodologia seguita dal Fact Checker soprattutto in merito ad alcune caratteristiche della **fn:EvidenceSource** dell'evidenza, formalizzate nel metodo **IMVAIN** (Figura 21 in appendice). Dal punto di vista della trasparenza, il Fact Checker dovrebbe riportare sia la sorgente da cui ottiene l'evidenza che una qualche risorsa deferenziale e accessibile (**fn:supportedBy**). In questo modo il lettore è in grado sia di accedere alla risorsa e analizzarla personalmente ma anche effettuare una serie di valutazioni sull'autorevolezza della fonte stessa. Per quanto riguarda l'autorevolezza della fonte, abbiamo modellato quest'ultima attraverso **fn:capacityType**, basandoci sulle linee guida proposte dall'associazione europea **EUFactcheck.eu**, in Figura 23 presente in appendice è riportato un esempio.
- fn:ClaimEvaluationActivity**: in questa fase uno o più Fact Checkers producono un articolo di **fn:Review** di un Claim, un esempio è riportato in Figura 4. Fatta eccezione per alcune realtà come **FactCheck.org**, la maggior parte delle organizzazioni utilizza un sistema di **fn:Rating** per valutare e indicare la "credibilità" del Claim. Il sistema utilizzato solitamente si basa su una scala ordinale, differente per ogni organizzazione. Spesso il valore di rating (**fn:ratingValue**) viene accompagnato da una sua rappresentazione multimediale (**fn:ratingAssociatedMedia**) e/o un breve testo (**fn:ratingSystemComment**) (Figura 1).

Elemento ricorrente in una Review è la presenza di una o più motivazioni (**fn:Judgment**) che spingono il Fact Checker ad assegnare un certo valore di rating [16], esse rappresentano una sorta di "verdetto riassuntivo" aiutando il lettore a cogliere rapidamente il contenuto della Review. Altro elemento ricorrente, è la presenza di **fn:mention** ad entità esterne (luoghi, persone, ecc...) in quanto fornisco informazioni di background che aiutano il lettore a contestualizzare il contenuto.



Figura 4: Esempio di Review con relativo Claim estratto da PagellaPolitica.com

Tassonomia delle motivazioni Abbiamo voluto modellare e sistematizzare le differenti motivazioni organizzandole in una tassonomia. La *motivation taxonomy* riportata in Figura 22, è stata ottenuta fondendo ed integrando due principali risorse: la [Washington Post Manipulated Video Taxonomy](#) (Figura 5) e le risorse educative messe a disposizione dal [European Journalism Training Association \(EJTA\)](#). La tassonomia proposta ha un duplice obiettivo: standardizzare il vocabolario usato dai Fact Checkers e strutturare l'informazione mettendo a disposizione annotazioni semantiche per gli strumenti di information retrieval.

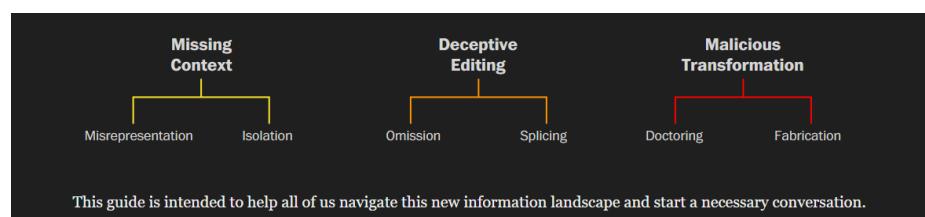


Figura 5: Estratto della Manipulated Video Taxonomy proposta dal [Washington Post](#).

3.1 Allineamento

La principale ontologia con cui abbiamo effettuato l'allineamento è stata [W3C PROV-O](#). L'ontologia PROV-O offre un modello flessibile per descrivere il flusso dei processi, gli input e gli output delle attività e l'attribuzione delle entità ai relativi agenti. Il livello di generalità e la nomenclatura offerta da PROV-O ha permesso un intuitivo processo di allineamento ontologico.

Il *Core Model* di PROV, riportato in Figura 6, è costituito da 3 componenti principali:

- *Entities*: rappresentano gli "oggetti" del processo, la cui natura può essere fisica, astratta o concettuale. La peculiarità del modello consiste nell'attribuire alle entità la provenienza.
- *Activities*: rappresentano gli aspetti dinamici del mondo che generano nuove entità o ne utilizzano di già esistenti.
- *Agents*: individui (persone, agenti software o organizzazioni) associati ad una attività che partecipano in essa con un certo grado di responsabilità ricoprendo un certo ruolo.

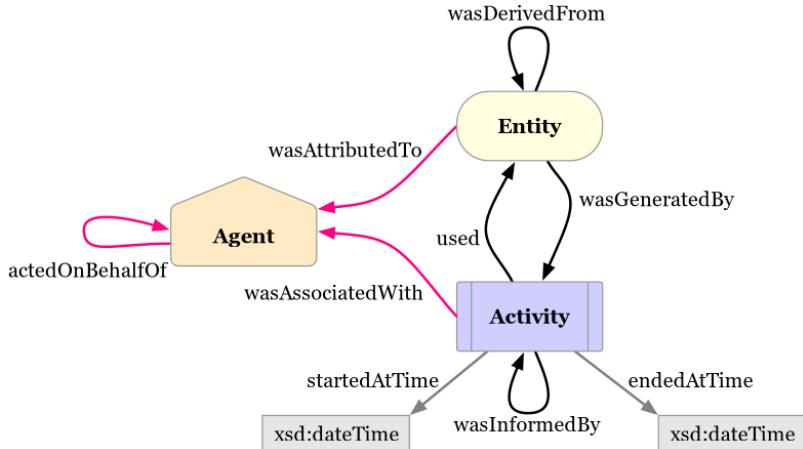


Figura 6: *Core Model* dell'ontologia PROV-O.

Data la generalità dei concetti del core model di PROV-O abbiamo allineato le entità di Fakt-News utilizzando la relazione di sussunzione attraverso il costrutto [rdfs:subClassOf](#). Come si può osservare in Figura 22, [prov:Activity](#), [prov:Agent](#), [prov:Entity](#) rappresentano le top-class di tutte le altre classi più specifiche del dominio.

Nel caso in cui i concetti espressi dalle entità presenti in Fakt-News fossero un "exact-match" con entità di ontologie ben più note, abbiamo deciso di utilizzare il costrutto [owl:EquivalentClass](#). Nello specifico abbiamo definito rispettivamente [fn:Person](#) e [fn:DigitalDocument](#) come classi definite in termini di [foaf:Person](#) e [schema:CreativeWork](#).

Ontology Design Patterns Nel rispetto delle best practices dell'*ontology engineering*, Fakt-News utilizza tre pattern ontologici. I primi due pattern, importati nell'ontologia dal repository [OntologyDesignPattern.org](#) sono stati **Set** e **TimeInterval**. L'utilizzo del pattern **Set** nasce dall'esigenza di dover descrivere [fn:AlternativesSet](#) come una collezione di elementi non duplicati. Mentre l'utilizzo del pattern **TimeInterval** nasce dall'esigenza di dover rappresentare gli istanti di inizio e la fine delle [prov:Activity](#) nel tempo.

Il terzo pattern utilizzato è il **Qualified Relations** definito dall'ontologia PROV-O. Questo pattern permette di qualificare una relazione, ovvero di poter descrivere una relazione specificandone

degli attributi. Il pattern è molto intuitivo, infatti per implementarlo è sufficiente definire la relazione come classe, creare degli individui e annotare la classe con le rispettive data e object properties. L'utilizzo di questo pattern è dovuto all'esigenza di dover rappresentare i **ruoli** dei vari **prov:Agent** nelle associazioni con le **prov:Activity**. PROV-O stessa utilizza questo pattern, estendendo il *Core Model* ed introducendo le Qualified Associations come mostrato in Figura 7.

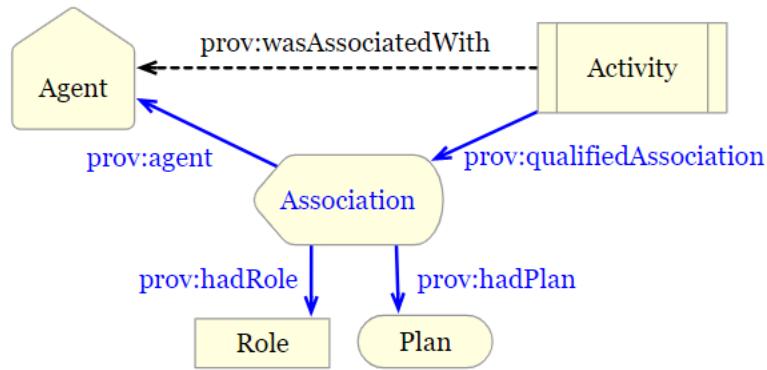


Figura 7: Diagramma del pattern *Qualified Association* implementato in PROV-O.

3.2 Knowledge Graph

In questa sezione riportiamo tre casi di principale interesse sulla *A-box* dell'ontologia.

Esempio 1: Vista Process-Centric Nell'esempio illustrato in Figura 8 mostriamo l'intero processo associato al Claim C-1 e Review R-1. Il focus è sul processo e si possono notare differenti caratteristiche: ogni Activity possiede una coppia di input-ouput, perdura nel tempo secondo un certo intervallo e "appartiene" allo stesso Plan.

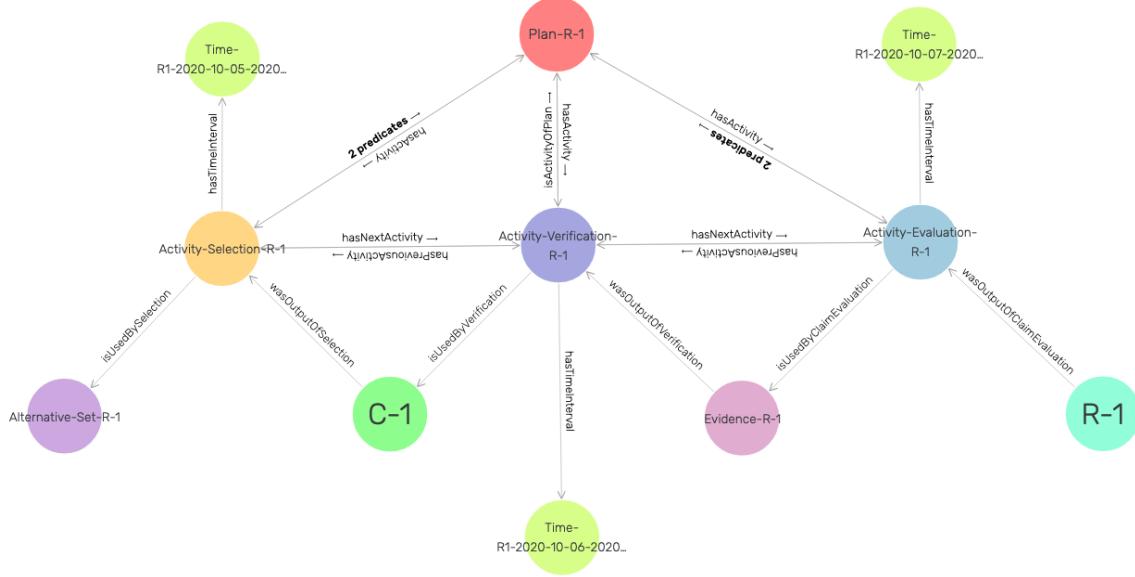


Figura 8: Rappresentazione a grafo dell'esempio 1.

	subject	predicate	object	context
1	fakt-news:Plan-R-1	fakt-news:hasActivity	fakt-news:Activity-Evaluation-R-1	http://www.ontotext.com/explicit
2	fakt-news:Plan-R-1	fakt-news:hasActivity	fakt-news:Activity-Selection-R-1	http://www.ontotext.com/explicit
3	fakt-news:Plan-R-1	fakt-news:hasActivity	fakt-news:Activity-Verification-R-1	http://www.ontotext.com/explicit
4	fakt-news:Plan-R-1	rdf:type	fakt-news:FactCheckingProcess	http://www.ontotext.com/explicit
5	fakt-news:Plan-R-1	rdf:type	owl:NamedIndividual	http://www.ontotext.com/explicit
6	fakt-news:Plan-R-1	rdf:type	prov:Entity	http://www.ontotext.com/explicit
7	fakt-news:Plan-R-1	rdf:type	prov:Plan	http://www.ontotext.com/explicit

Figura 9: Rappresentazione a triple RDF dell'esempio 1. **Nota:** non sono presenti tutte le triple per ragioni di leggibilità, ma solamente le triple principali che caratterizzano l'istanza Plan-R-1.

Esempio 2: Vista Review-Centric Nell'esempio illustrato in Figura 10 mostriamo le principali relazioni tra la Review R-1 e le principali entità ad essa associate. Nella Figura 11 sono riportate le triple RDF di cui è composta l'istanza R-1.

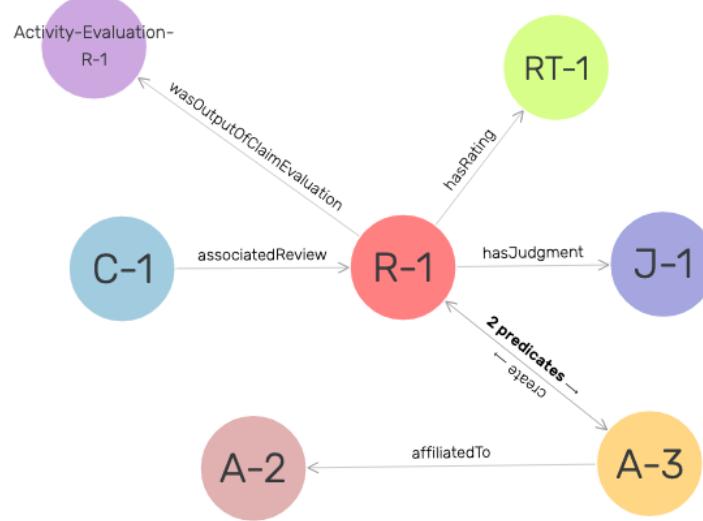
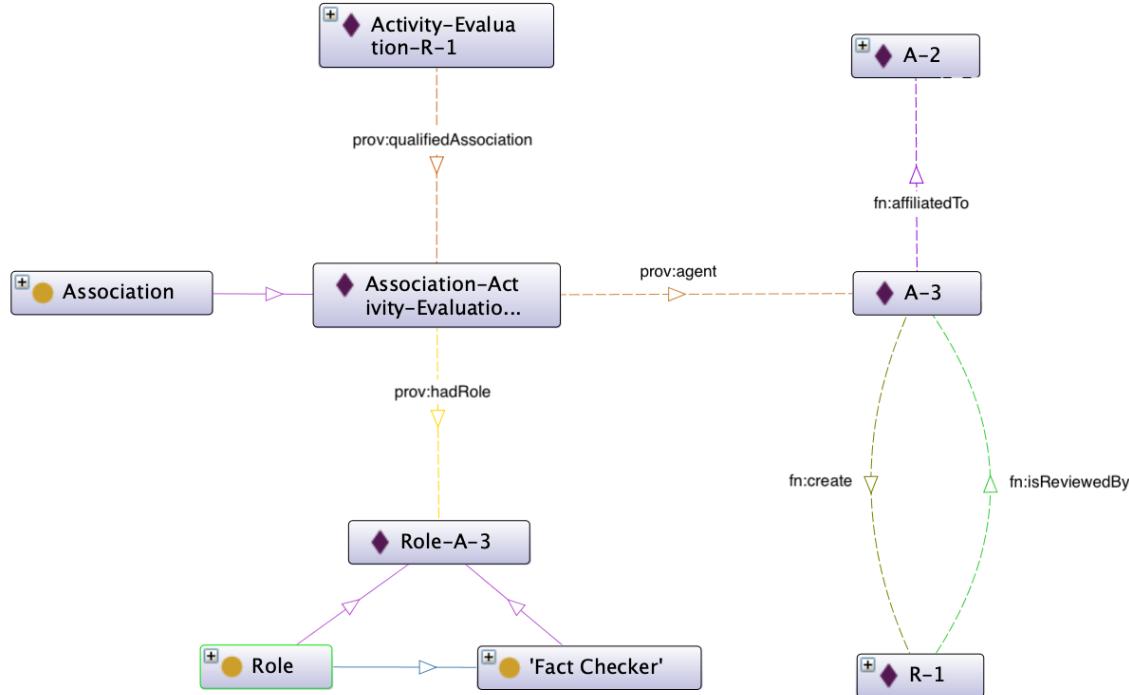


Figura 10: Rappresentazione a grafo dell'esempio 2.

	subject	predicate	object	context
1	fakt-news-R-1	terms:date	"2020-10-08"^^xsd:date	http://www.ontotext.com/explicit
2	fakt-news-R-1	terms:language	"it"	http://www.ontotext.com/explicit
3	fakt-news-R-1	terms:title	"Il governo ha davvero abolito i 'decreti sicurezza'?"	http://www.ontotext.com/explicit
4	fakt-news-R-1	schema:url	"https://www.avocca.info/archives/69816/il-governo-ha-davvero-abolito-i-decreti-sicurezza?#xsevqj0"	http://www.ontotext.com/explicit
5	fakt-news-R-1	fakt-news:creator	fakt-news:A-3	http://www.ontotext.com/explicit
6	fakt-news-R-1	fakt-news:documentContent	"Il fact-checking de lavice.info passa ...+[+1083 word s]"	http://www.ontotext.com/explicit
7	fakt-news-R-1	fakt-news:isJudgment	fakt-news:J-1	http://www.ontotext.com/explicit
8	fakt-news-R-1	fakt-news:isHastating	fakt-news:RT-1	http://www.ontotext.com/explicit
9	fakt-news-R-1	fakt-news:isReviewedBy	fakt-news:A-3	http://www.ontotext.com/explicit
10	fakt-news-R-1	fakt-news:mention	"Nicola_Zingaretti"	http://www.ontotext.com/explicit
11	fakt-news-R-1	fakt-news:reviewContent	"Il fact-checking de lavice.info passa ...+[+1083 word s]"	http://www.ontotext.com/explicit
12	fakt-news-R-1	fakt-news:tag	"politica"	http://www.ontotext.com/explicit
13	fakt-news-R-1	fakt-news:isOutputOfClaimEvaluation	fakt-news:Activity-Evaluation-R-1	http://www.ontotext.com/explicit
14	fakt-news-R-1	rdf:type	schema:CreativeWork	http://www.ontotext.com/explicit
15	fakt-news-R-1	rdf:type	fakt-news:DigitalDocument	http://www.ontotext.com/explicit
16	fakt-news-R-1	rdf:type	fakt-news:Review	http://www.ontotext.com/explicit
17	fakt-news-R-1	rdf:type	owl:NamedIndividual	http://www.ontotext.com/explicit
18	fakt-news-R-1	rdf:type	prov:Entity	http://www.ontotext.com/explicit
19	fakt-news-R-1	prov:wasAttributedTo	fakt-news:A-3	http://www.ontotext.com/explicit
20	fakt-news-R-1	prov:wasGeneratedBy	fakt-news:Activity-Evaluation-R-1	http://www.ontotext.com/explicit

Figura 11: Rappresentazione a triple RDF dell'esempio 2.

Esempio 3: Qualified Association Nell'esempio illustrato in Figura 12 è illustrata una Qualified Association tra la Review R-1 e Activity-Evaluation-R-1. In questo esempio si può apprezzare l'utilizzo del pattern *Qualified Relation* che ha permesso di specificare il ruolo di Fact Checker dell'agente A-3.



	subject	predicate	object	context
1	fakt-news:Association-Activity-Evaluation-R1-A3	rdf:type	owl:NamedIndividual	http://www.ontotext.com/explicit
2	fakt-news:Association-Activity-Evaluation-R1-A3	rdf:type	prov:Association	http://www.ontotext.com/explicit
3	fakt-news:Association-Activity-Evaluation-R1-A3	prov:agent	fakt-news:A-3	http://www.ontotext.com/explicit
4	fakt-news:Association-Activity-Evaluation-R1-A3	prov:hadRole	fakt-news:Role-A-3	http://www.ontotext.com/explicit

Figura 13: Rappresentazione in triple RDF dell'istanza **Association-Activity-Evaluation-R1-A3**. Anche qui come nell'Esempio 1 non sono state riportate tutte le triple che caratterizzano la Qualified Association per ragioni di leggibilità.

3.3 Reasoning

Di seguito riportiamo le principali inferenze di maggiore interesse.

Relazione di Sussunzione Avendo definito `fn:Author` come classe definita il reasoner sussume che `fn:Factchecker` e `fn:Claimant` sono sottoclassi di `fn:Author`. Per le stesse motivazioni, `fn:Claim` e `fn:Review` vengono classificate come sottoclassi di `schema:CreativeWork`. Queste inferenze sono visibili nella Figura 22 in appendice.

Fact Checking Process Afferendo gli individui di classe `fn:Plan`, `fn:SelectionActivity`, `VerificationActivity` e `fn:ClaimEvaluationActivity`, con loro i rispettivi input e output, e avendo definito `fn:hasNextActivity` come proprietà transitiva e inversa di `fn:hasPreviousActivity`, e avendo asserito `fn:isFirstActivityOfPlan` e `fn:isLastActivityOfPlan` il reasoner è in grado di effettuare la **chiusura transitiva** delle rispettive relazioni. In seguito a queste inferenze, viene inferito che l'individuo *Plan* è di tipo `fn:FactCheckingProcess`, poiché esso risulta equivalente a un `prov:Plan` in relazione con le tre differenti `prov:Activity`.

Property Chain Sono state realizzate due property chains:

- `fn:claimReviewed`: permette di inferire il Claim associato ad una Review attraverso la seguente inferenza, "se una Review è l'output di una attività di Claim evaluation e l'attività che la precede utilizza un Claim \vdash l'articolo è una Review per quel Claim". In notazione Manchester:

wasOutputOfClaimEvaluation o hasPreviousActivity o prov:used **SubPropertyOF**
claimReviewed

- `fn:isCertifiedBy`: permette di inferire se un'organizzazione di fact checking è certificata da una qualche autorità. Ovvero "se un'organizzazione segue dei un codice di condotta creato da un'autorità di fact checking \vdash l'organizzazione è certificata". In notazione Manchester:

observeCodeOfPrinciples o creator **SubPropertyOF** isCertifiedBy

In Figura 14 sono riportate le property chain in modo visuale.

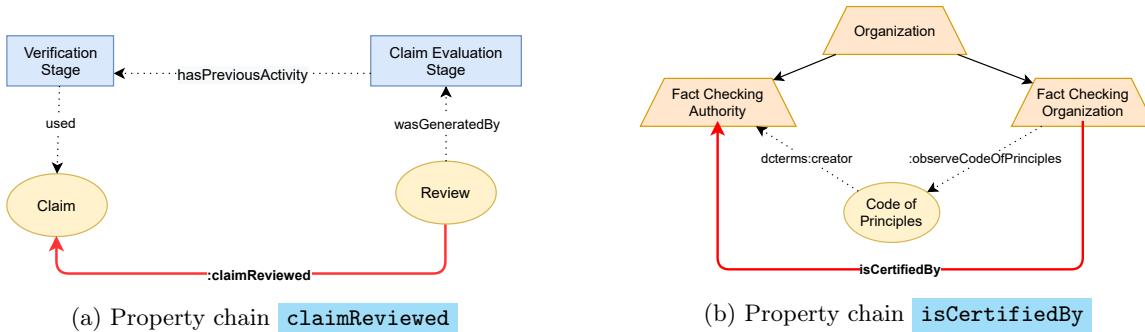


Figura 14: Rappresentazione grafica delle due property chain (in rosso).

4 Piattaforma LOD

Come RDF store è stato scelto *GraphDB 9.6.0 Free*. Il file `create_rdf_store.py` contiene la configurazione della repository, tra cui l'id della repository: `modsem-faktnews` e il prefisso base dell'ontologia: <http://www.modsem.org/faktnews>. La creazione del repository e il caricamento dei dati avviene in modo programmatico sfruttando le REST API messe a disposizione da GraphDB. In Figura 15 è rappresentato l'intero stack software utilizzato.

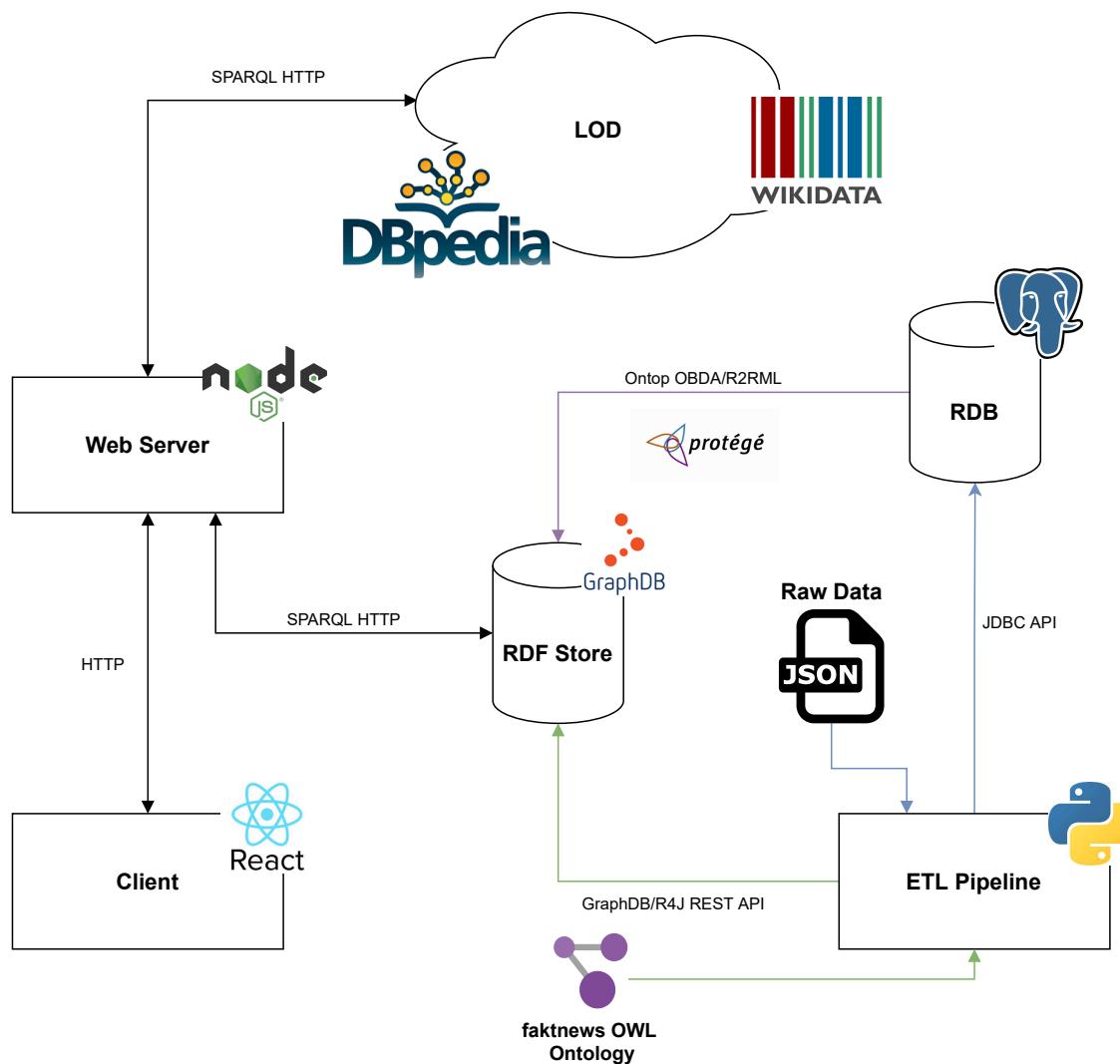


Figura 15: Diagramma delle principali componenti software e dei protocolli di comunicazione utilizzati

4.1 Queries SPARQL

In questa sezione riportiamo le queries create a supporto dell’interazione con l’applicazione (Figura 16). Per semplicità di lettura, elenchiamo i prefissi comuni a tutte le queries. Quelli riportati nelle singole query sono da intendersi come specifici della query in questione. Tutte le query (file `queries.sparql`), e di conseguenza i *result sets* in output, sono state eseguite sul knowledge graph ottenuto dalla **materializzazione** delle inferenze del reasoner *Pellet*.

```

PREFIX : <http://www.modsem.org/fakt-news#>
PREFIX fn: <http://www.modsem.org/fakt-news#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

```

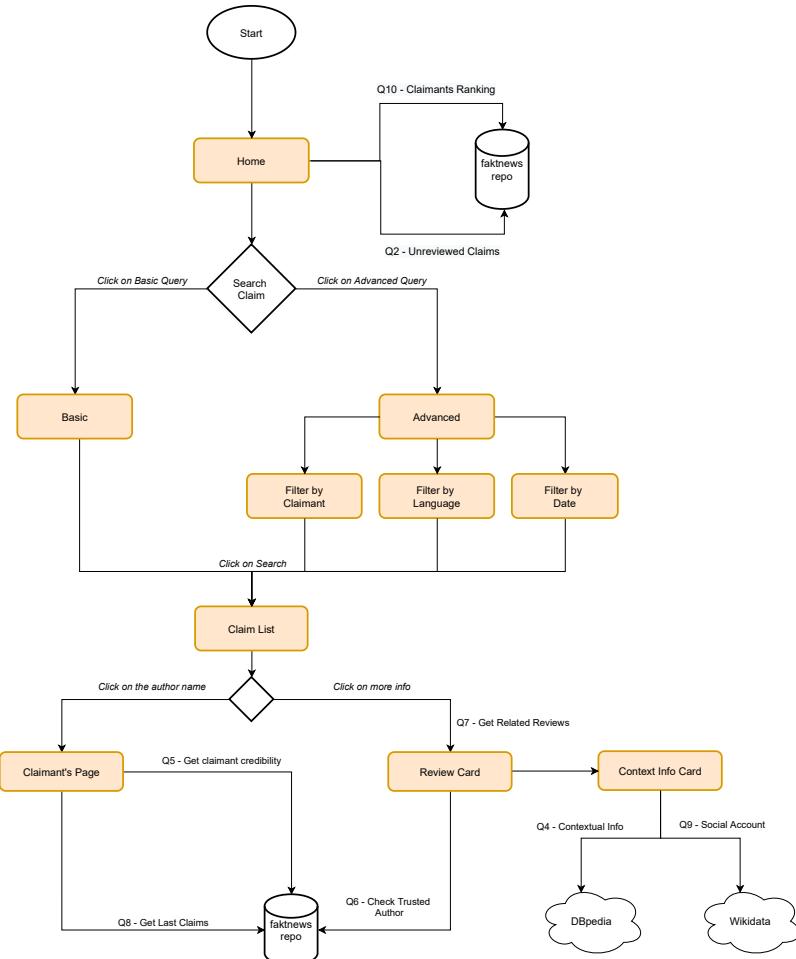


Figura 16: Diagramma dell’interazione con l’applicazione. I riquadri evidenziati in colore arancione rappresentano le componenti della UI

Q1 Ricerca Reviews per Keyword La seguente query permette di recuperare tutte le reviews che contengono sia nel titolo che nel corpo del testo una data keyword. La ricerca è di tipo "approssimato", ovvero viene utilizzata un'espressione regolare con la funzione REGEX. Si osservi l'utilizzo del connettivo logico or (||) nel costrutto FILTER.

La query può essere facilmente parametrizzata, in questo caso specifico immaginiamo che l'utente scriva in input "video".

```
PREFIX schema: <http://schema.org/>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT ?fc_name ?review_title ?review_text
WHERE {
    ?review a fn:Review;
        fn:isReviewedBy ?factchecker;
        dct:title ?review_title;
        fn:documentContent ?review_text .
    ?factchecker schema:name ?fc_name .

    BIND ("video" as ?keyword)
    FILTER(REGEX(?review_title, ?keyword, "i") ||
           REGEX(?review_text, ?keyword, "i")) .
} LIMIT 100
```

Results set della query.

	fc_name	review_title	review_text
1	"Allyson Chiu"	"Facebook wouldn't delete an altered video of Nancy Pelosi. What about one of Mark Zuckerberg?"	"A couple of short video clips, a voice actor and an artificial intelligence algorithm. That's all two artists and a technology start-up ...[+1497 words]"
2	"Daniel Funke"	"Barack Obama wasn't arrested before Biden's inauguration"	"In a video published Jan. 17, Gage Nelson, host of the conservative podcast America Divided, said: "Barack Obama was arrested."'

Q2 Claims non Verificati La query permette di recuperare i claims presenti nel KG non revisionati da nessun fact-checker. Vengono recuperate le reviews associate ad un claim e utilizzando il costrutto MINUS vengono rimosse dal results set. In questo modo verranno recuperati solo i claims senza una review associata.

```
PREFIX schema: <http://schema.org/>

SELECT ?claim ?claimant_name
WHERE {
    ?claim a :Claim; :isClaimedBy ?claimant .
    ?claimant schema:name ?claimant_name .
    MINUS {
        ?rev :claimReviewed ?claim .
    }
} LIMIT 100
```

Il result set della query è vuoto in quanto nel KG ogni claim è associato ad almeno una review.

Q3 Filtro Temporale La seguente query permette di recuperare tutti i claims presenti nel KG dichiarati in un dato intervallo temporale. La particolarità della query consiste nell'utilizzo del costruttore xsd:datetime in quanto il confronto tra date è specificato dallo standard solo per tipo xsd:datetime mentre ?claim_date è di tipo xsd:date.

La query può essere facilmente parametrizzata, nel caso specifico vengono recuperati tutti i claims dichiarati nell'anno tra il 2019 e il 2020.

```

PREFIX schema: <http://schema.org/>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT ?title ?claim_date
WHERE {
  ?claim a :Review; dct:date ?claim_date; dct:title ?title
  BIND("2019-01-01T00:00:00Z"^^xsd:dateTime AS ?first_date)
  BIND("2020-12-01T00:00:00Z"^^xsd:dateTime AS ?last_date)
  BIND(xsd:dateTime(?claim_date) AS ?claim_datetime)
  FILTER( ?claim_datetime > ?first_date && ?claim_datetime < ?last_date )
}

} limit 100

```

Result set della query:

	title	claim_date
1	"Il governo ha davvero abolito i "decreti sicurezza"?"	"2020-10-08"^^xsd:date
2	"Facebook wouldn't delete an altered video of Nancy Pelosi. What about one of Mark Zuckerberg?"	"2019-06-19"^^xsd:date

Q4 Informazioni Contestuali Query federata. Nella prima parte della query, vengono estratte le entità menzionate nelle review. Da notare l'utilizzo del costrutto **OPTIONAL** in quanto nella maggior parte dei casi le entità menzionate non sono sempre presenti.

Nella seconda parte viene utilizzato il costrutto **SERVICE** per effettuare una query all'endpoint pubblico di *DBpedia*. Di particolare importanza sono i comandi utilizzati per il bind della variabile **?entity**. L'uri della risorsa di DBpedia viene costruito **dinamicamente** dal prefisso **dbr:** e dalla variabile **?wiki_entity**, poichè nella nostra ontologia codifica il **Page_name** delle pagine di *Wikipedia*.

Dato l'aspetto multilingua di *DBpedia*, il costrutto **FILTER** è stato utilizzato per estrarre solo le risorse con uno specifico language tag (in questo caso "en").

La query federata estrae alcune informazioni di background sulle entità menzionate:

- **?entity_label**: label associata alla risorsa.
- **?info**: commento/abstract associato alla risorsa.
- **?linked_res_wiki**: url delle pagine wikipedia di risorse secondarie strettamente correlate a quella menzionata.

```

PREFIX dct: <http://purl.org/dc/terms/>
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?review_title ?entity_label ?info ?linked_res_wiki
WHERE {
    ?review a :Review;
            dct:title ?review_title.
    OPTIONAL {?review :mention ?wiki_entity.}
    BIND(URI(CONCAT(STR(dbr:), STR(?wiki_entity))) AS ?entity).

    SERVICE <https://dbpedia.org/sparql> {
        ?entity rdfs:label ?entity_label.
        OPTIONAL {?entity rdfs:seeAlso ?linked_res.
                  ?linked_res foaf:isPrimaryTopicOf ?linked_res_wiki}.
        OPTIONAL {?entity rdfs:comment ?info}.
        FILTER(langMatches(lang(?entity_label), "en") &&
              langMatches(lang(?info),"en"))
    }
} limit 100

```

Result set della query:

review_title	entity_label	info	linked_res_wiki
Barack Obama wasn't arrested before Biden's inauguration	QAnon	QAnon () is a far-right conspiracy theory detailing a supposed secret plot by an alleged "deep state" against U.S. President Donald Trump and his supporters. The theory began with an October 2017 post on the anonymous imageboard 4chan by someone using the name Q, who was presumably an American individual initially, but probably later became a group of people, claiming to have access to classified information involving the Trump administration and its opponents in the United States. Analysis by NBC News found that three people took the original Q post and expanded it across multiple media platforms to build internet followings for monetization. Qanon was preceded by several similar anonymous 4chan posters such as FBIAnon, HLIAnon (high level insider), CIAAnon and WH Insider Anon.	http://en.wikipedia.org/wiki/Murder_of_Seth_Rich
Barack Obama wasn't arrested before Biden's inauguration	QAnon	QAnon () is a far-right conspiracy theory detailing a supposed secret plot by an alleged "deep state" against U.S. President Donald Trump and his supporters. The theory began with an October 2017 post on the anonymous imageboard 4chan by someone using the name Q, who was presumably an American individual initially, but probably later became a group of people, claiming to have access to classified information involving the Trump administration and its opponents in the United States. Analysis by NBC News found that three people took the original Q post and expanded it across multiple media platforms to build internet followings for monetization. Qanon was preceded by several similar anonymous 4chan posters such as FBIAnon, HLIAnon (high level insider), CIAAnon and WH Insider Anon.	http://en.wikipedia.org/wiki/Pizzagate_conspiracy_theory
Barack Obama wasn't arrested before Biden's inauguration	Barack Obama	Barack Hussein Obama I () (born August 4, 1961) is an American politician and attorney who served as the 44th president of the United States from 2009 to 2017. A member of the Democratic Party, Barack Obama was the first African-American president of the United States. He previously served as a U.S. senator from Illinois from 2005 to 2008 and an Illinois state senator from 1997 to 2004.	http://en.wikipedia.org/wiki/Climate_change_policy_of_the_United_States
Barack Obama wasn't arrested before Biden's inauguration	Barack Obama	Barack Hussein Obama I () (born August 4, 1961) is an American politician and attorney who served as the 44th president of the United States from 2009 to 2017. A member of the Democratic Party, Barack Obama was the first African-American president of the United States. He previously served as a U.S. senator from Illinois from 2005 to 2008 and an Illinois state senator from 1997 to 2004.	http://en.wikipedia.org/wiki/Russia-United_States_relations
Barack Obama wasn't arrested before Biden's inauguration	Barack Obama	Barack Hussein Obama II () (born August 4, 1961) is an American politician and attorney who served as the 44th president of the United States from 2009 to 2017. A member of the Democratic Party, Barack Obama was the first African-American president of the United States. He previously served as a U.S. senator from Illinois from 2005 to 2008 and an Illinois state senator from 1997 to 2004.	http://en.wikipedia.org/wiki/The_Syrian_Civil_War
Barack Obama wasn't arrested before Biden's	Barack Obama	Barack Hussein Obama II () (born August 4, 1961) is an American politician and attorney who served as the 44th president of the United States from 2009 to 2017. A member of the Democratic Party, Barack Obama was the first African-American president of the United States. He previously served as a U.S. senator from Illinois from 2005 to 2008 and an Illinois state senator from 1997 to 2004.	http://en.wikipedia.org/wiki/Confirmations_of_Barack_Obama's_Cabinet

Q5 Credibilità di un Claimant La seguente query permette di "verificare la credibilità" (in modo approssimato) di un individuo. Nello specifico viene effettuata un operazione di aggregazione e successivamente un conteggio di quanti dei claim dichiarati siano stati valutati "true", ..., "false".

```
PREFIX schema: <http://schema.org/>

SELECT ?claimant_name ?rating_val (COUNT(?review) as ?n_reviews)
WHERE {
    ?review rdf:type :Review;
        :hasRating ?rating;
        :claimReviewed ?claim .
    ?claim :isClaimedBy ?claimant .
    ?claimant schema:name ?claimant_name .
    ?rating :ratingValue ?rating_val .
}
GROUP BY ?claimant_name ?rating_val
ORDER BY ?claimant_name ?rating_val
```

Result set della query:

	claimant_name	rating_val	n_reviews
1	"America Divided Podcast"	"False"	"1"^^xsd:integer
2	"Mark Zuckerberg"	"False"	"1"^^xsd:integer
3	"Matteo Salvini"	"Mostly False"	"1"^^xsd:integer
4	"Nicola Zingaretti"	"True"	"1"^^xsd:integer

Q6 Autore Verificato La seguente query permette di verificare che l'autore della review appartenga a una organizzazione di Fact Checking verificata, ovvero che rispetta i principi fondamentali di una qualche autorità. Viene utilizzato il costrutto **ASK** in quanto la query vuole solo verificare se un determinato autore è verificato o meno. Inoltre, notare come il costrutto **BIND** permetta una parametrizzazione della query.

Il costrutto **UNION** viene utilizzato per differenziare due casi possibili. Infatti è possibile che l'autore di una review sia un qualche agente affiliato ad un'organizzazione di fact-checking o l'organizzazione stessa.

```

PREFIX schema: <http://schema.org/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX prov: <http://www.w3.org/ns/prov#>

ASK {
    BIND(URI(concat(str(fn:), "A-8")) as ?reviewer)
    ?reviewer fn:create ?review.
    {
        ?reviewer a fn:Organization.
        ?reviewer fn:observeCodeOfPrinciples ?code.
    } UNION {
        ?reviewer fn:affiliatedTo ?fco.
        ?fco a prov:Agent; fn:observeCodeOfPrinciples ?code.
        ?fco schema:name ?fcoName.
    }
}

```

Dato l'utilizzo di **ASK**, la query non restituisce un result set ma una risposta binaria "yes"/"no", in questo caso per l'agente *A-8* restituisce "yes".

Q7 Reviews associate a uno specifico Claim Poiché è possibile che per un claim esistano più di una review, la seguente query restituisce la lista di tutte le review associate ad uno specifico claim. La query può essere facilmente parametrizzata. Per ogni review vengono listati alcuni dettagli, come il *rating*, il *judgment*, il *content*, il *language*, il *tag*, l'*URL* e le *mentions*. Da notare l'utilizzo del costrutto **OPTIONAL** in quanto spesso le review non contengono tags o entità menzionate.

```

PREFIX schema: <http://schema.org/>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT ?claim ?reviewer ?rating ?judgment ?content ?language ?url
      ?tag ?mention
WHERE {
    ?claim a fn:Claim; fn:associatedReview ?associatedReview .
    ?agent schema:name ?reviewer .
    ?associatedReview fn:isReviewedBy ?agent ;
                    fn:hasRating ?r ;
                    fn:hasJudgment ?j ;
                    fn:reviewContent ?content ;
                    dct:language ?language ;
                    schema:url ?url .
}
```

```

?r fn:ratingValue ?rating .
?j fn:judgmentSummary ?judgment .
OPTIONAL{ ?associatedReview fn:tag ?tag;
fn:mention ?mention.}
} LIMIT 100

```

Result set della query:

claim	reviewer	rating	Judgment	content	language	url	tag	mention
fakt-news-C-2	Pagella Politica	Mostly False	Missing Context	Al governo si preoccupano di cancellare "padre" e "madre" dalla carta di identità dei minori per sostituirli con genitore 1 e 2 ...[+1568 words]	it	https://pagellapolitica.it/dichiarazioni/8801/no-genitore-1-e-2-non-tornano-sui-dокументi-identita-non-ci-sono-mai-stati	politica	Matteo_Salvini
fakt-news-C-2	Pagella Politica	Mostly False	Missing Context	Al governo si preoccupano di cancellare "padre" e "madre" dalla carta di identità dei minori per sostituirli con genitore 1 e 2 ...[+1568 words]	it	https://pagellapolitica.it/dichiarazioni/8801/no-genitore-1-e-2-non-tornano-sui-dокументi-identita-non-ci-sono-mai-stati	Carta d'identità	Matteo_Salvini
fakt-news-C-2	Pagella Politica	Mostly False	Missing Context	Al governo si preoccupano di cancellare "padre" e "madre" dalla carta di identità dei minori per sostituirli con genitore 1 e 2 ...[+1568 words]	it	https://pagellapolitica.it/dichiarazioni/8801/no-genitore-1-e-2-non-tornano-sui-dокументi-identita-non-ci-sono-mai-stati	Lega	Matteo_Salvini
fakt-news-C-3	Allison Chiu	False	Fabrication	A couple of short video clips, a voice actor and an artificial intelligence algorithm. That's all two artists and a technology start-up ...[+1497 words]	en	https://www.washingtonpost.com/nation/2019/06/12/mark-zuckerberg-deepfake-facebook-instagram-nancy-pelosi/	Nancy_Pelosi	Mark_Zuckerberg
fakt-news-C-3	Allison Chiu	False	Fabrication	A couple of short video clips, a voice actor and an artificial intelligence algorithm. That's all two artists and a technology start-up ...[+1497 words]	en	https://www.washingtonpost.com/nation/2019/06/12/mark-zuckerberg-deepfake-facebook-instagram-nancy-pelosi/	deepfake	Mark_Zuckerberg
fakt-news-C-4	Daniel Funke	False	Deceptive Editing	In a video published Jan. 17, Gage Nelson, host of the conservative podcast America Divided, said: "Barack Obama was arrested."	en	https://www.politifact.com/factchecks/2021/jan/18/facebook-posts/obama-wasn't-arrested-bidens-inauguration/	politics	Barack_Obama
fakt-news-C-4	Daniel Funke	False	Deceptive Editing	In a video published Jan. 17, Gage Nelson, host of the conservative podcast America Divided, said: "Barack Obama was arrested."	en	https://www.politifact.com/factchecks/2021/jan/18/facebook-posts/obama-wasn't-arrested-bidens-inauguration/	politics	Nelson_Mandela
fakt-news-C-4	Daniel Funke	False	Deceptive Editing	In a video published Jan. 17, Gage Nelson, host of the conservative podcast America Divided, said: "Barack Obama was arrested."	en	https://www.politifact.com/factchecks/2021/jan/18/facebook-posts/obama-wasn't-arrested-bidens-inauguration/	politics	QAnon
fakt-news-C-1	Massimo Taddei	True	Confirmed Fact	Il fact-checking de lavoce.info passa ...[+1283 words]	it	https://www.lavoice.info/archives/69816/il-governo-ha-davvero-abolito-i-decreti-sicurezza/	politica	Nicola_Zingaretti

Q8 Ultimi Claims di un Claimant La seguente query restituisce la lista degli ultimi K claims di un dato claimant. Per ogni claim, vengono listati il *claimant*, l'*URL*, il *content* e il *language*. Nel caso specifico $K = 5$ e il claimant è *A-11*.

```

PREFIX schema: <http://schema.org/>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT ?claimantName ?claim ?url ?content
WHERE {
  BIND(URI(concat(str(fn:), "A-11")) as ?claimant).

  ?claimant schema:name ?claimantName .
  ?claim a fn:Claim;
         fn:isClaimedBy ?claimant;
         dct:date ?date;
         schema:url ?url;
         fn:claimContent ?content.
}
ORDER BY DESC(?date)
LIMIT 5

```

Result set della query:

	claimantName	claim	url	content
1	"Mark Zuckerberg"	fakt-news:C-3	"https://www.instagram.com/p/ByaVigGFP2U/?utm_source=ig_embed&utm_campaign=embed_video&watch_again">xsd:anyURI	Zuckerberg: we are increasing transparency on Ads. Announces new measure to protect the elections."

Q9 Social Network Account La seguente query permette di recuperare l'account dei principali social network delle entità menzionate nelle reviews. La query utilizza il meccanismo di federazione utilizzando due endpoint pubblici.

Nella prima parte vengono estratte le entità menzionate nelle reviews. Successivamente viene inviata una richiesta all'endpoint di DBpedia per recuperare l'uri della stessa entità in Wikidata. Questo è possibile grazie a `owl:sameAs` presente nella maggiorparte delle risorse di DBpedia e una successiva operazione di filtraggio per eliminare tutti gli uri non associati a Wikidata.

Una volta estratto il codice dell'entità (es. *Q76* per Barack Obama), viene costruito l'uri del soggetto, informazione contenuta in `?wiki_entity`, da utilizzare nella seconda query federata all'endpoint di Wikidata.

```

PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wd: <http://www.wikidata.org/entity/>

SELECT DISTINCT ?entity_label ?twt ?fb ?ig
WHERE {
    ?review a :Review;
    OPTIONAL {?review :mention ?entity.}
    BIND(URI(CONCAT(STR(dbr:), STR(?entity))) AS ?dbp_entity).

    SERVICE <https://dbpedia.org/sparql> {
        ?dbp_entity a foaf:Person;
                    rdfs:label ?entity_label
        OPTIONAL{?dbp_entity owl:sameAs ?linked_res} repositories

        FILTER(contains(str(?linked_res), "wikidata"))
        BIND(REPLACE(str(?linked_res), "(.+?)([A-Z0-9]+)", "$2") AS ?wiki_code)
        BIND(URI(CONCAT(STR(wd:), STR(?wiki_code))) AS ?wiki_entity)
    }
    FILTER(langMatches(lang(?entity_label), "en"))

    SERVICE <https://query.wikidata.org/sparql> {
        OPTIONAL{?wiki_entity wdt:P2013 ?fb_name;
                  wdt:P2002 ?twt_name.}
        OPTIONAL{?wiki_entity wdt:P2003 ?ig_name.}

        BIND(URI(CONCAT("https://www.facebook.com/", ?fb_name)) as ?fb)
        BIND(URI(CONCAT("https://twitter.com/", ?twt_name)) as ?twt)
        BIND(URI(CONCAT("https://www.instagram.com/", ?ig_name)) as ?ig)
    }
}

```

Result set della query:

	entity_label	tw	fb	ig
1	"Barack Obama"@en	https://twitter.com/BarackObama	https://www.facebook.com/barackobama	https://www.instagram.com/barackobama
2	"Nelson Mandela"@en			
3	"Nicola Zingaretti"@en	https://twitter.com/nzingaretti	https://www.facebook.com/nicolazingaretti	https://www.instagram.com/nicolazingaretti
4	"Matteo Salvini"@en	https://twitter.com/matteosalvinimi	https://www.facebook.com/salviniofficial	https://www.instagram.com/matteosalviniofficial
5	"Mark Zuckerberg"@en	https://twitter.com/finkd	https://www.facebook.com/zuck	https://www.instagram.com/zuck

Q10 Ranking dei Claimants più Attivi La seguente query restituisce la classifica dei claimants in base al numero di claim effettuati da questi ultimi. Tale classifica è descrescente, come si può notare dall'utilizzo del costrutto ORDER BY DESC. Nello specifico, la query restituisce in output il nome del claimant e il numero di claim associati grazie ad un operazione di aggregazione e l'uso del costrutto COUNT.

```
PREFIX schema: <http://schema.org/>

SELECT ?claimant (count(?claim) as ?numberClaims)
WHERE {
    ?claim a :Claim; :isClaimedBy ?someone .
    ?someone schema:name ?claimant .
}
GROUP BY ?claimant
ORDER BY DESC(?numberClaims)
LIMIT 100
```

Result set della query:

	claimant	numberClaims
1	"America Divided Podcast"	*1* xsd:integer
2	"Mark Zuckerberg"	*1* xsd:integer
3	"Matteo Salvini"	*1* xsd:integer
4	"Nicola Zingaretti"	*1* xsd:integer

5 Mapping R2RML

Dataset Al fine di creare un KG il più rappresentativo possibile abbiamo deciso di utilizzare come sorgenti di dati due dataset: *LiarPlus* [17] e *MultiFC* [18]. Entrambi i dataset sono stati costruiti effettuando web scraping dai principali siti di fact checking in lingua inglese.

Successivamente ad una prima fase di esplorazione dei dati, è stato necessario definire una pipeline ETL (Extract-Transform-Load) per pulire i dati (normalizzazione e imputing), ristrutturarli in base allo schema relazionale del database (Figura 17) e caricarli nel database. Per ulteriori informazioni si rimanda alla lettura dei commenti presenti nel codice (`etl_pipeline.py`, `cleaning_pipeline.py`).

Database Per semplificare lo schema relazionale, abbiamo deciso di concentrarci sulla "vista" dell'ontologia *review-centric*, ovvero le entità mappate riguardano principalmente le classi, object properties e data properties in relazione con individui di classe Review. Le entità relative alla vista *process-centric* sono state escluse nella modellazione dello schema relazionale (Figura 17) e dal mapping R2RML. Questa scelta è stata dettata principalmente dalla mancanza di sorgenti di dati che descrivono il processo di fact-checking. Per quanto riguarda lo stack software è stato utilizzato *PostgreSQL 13.1* come RDBMS e la piattaforma *OnTop* per il mapping R2RML (Figura 15).

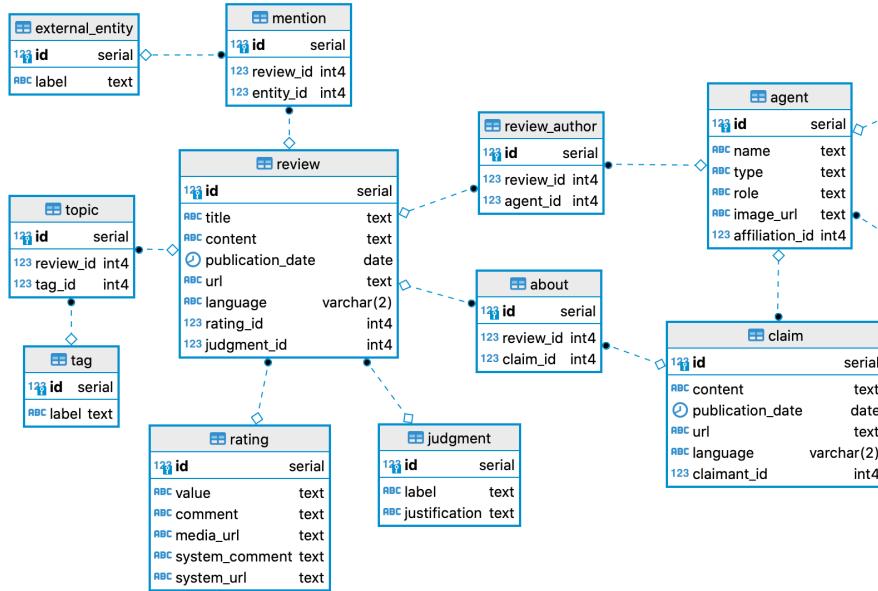


Figura 17: Lo schema relazionale del database.

OnTop Al fine di effettuare il mapping R2RML è stato utilizzata la piattaforma OnTop (in versione plugin per Protégé), che grazie all'opportuno driver JDBC permette a Protégé di interfacciarsi con i dati presenti nel RBDMS. Tramite le regole di mapping R2RML è stato effettuato il popolamento dei principali individui dell'ontologia.

Di seguito viene riportato un esempio di mapping sulla classe *Claim*. Il codice è stato diviso per favorire la leggibilità, in Tabella 1 viene illustrato il mapping così come viene formattato in Protégé, mentre nel listato 3 (visionabile in appendice) il corrispettivo mapping in linguaggio R2RML.

mappingId MAPID-Claim	
source	:C-{cid} a :Claim ; :associatedReview :R-{rid} ; :isClaimedBy :A-{aid} ; :claimContent {ccn} ; dct:date {cpd} ; schema:url {curl}^^xsd:anyURI ; dct:language {cla} .
target	<pre> SELECT claim.id AS cid, claim.content AS ccn, claim.publication_date AS cpd, claim.url AS curl, claim.language AS cla, claim.claimant_id AS aid, review.id AS rid FROM claim, about, review WHERE review.id = about.review_id AND about.claim_id = claim.id ; </pre>

Tabella 1: Mapping sulla classe Claim.

È importante notare che al momento della materializzazione delle triple, OnTop non solo importa i dati dal RDBMS seguendo i mapping R2RML, ma arricchisce tali dati con delle inferenze prodotte dal proprio Reasoner. Tali inferenze riguardano le relazioni di sussunzione (*subclassOf*) e relazioni inverse, mentre risultano assenti le proprietà più complesse come le *property chain*. Infatti, per asserire le proprietà mancanti è stato necessario avviare un secondo Reasoner (*Pellet*).

Di seguito riportiamo un esempio di come un singolo record relazionale (ottenuto dalla query SQL in Tabella 3) viene mappato in un individuo (Figura 18).

cid	ccn	cpd	curl	cla	aid	rid
0	We ended up get...	2017-01-31	https://www.factcheck.org ... en	31	0	

Tabella 2: Esempio di record relazionale oggetto del mapping

Figura 18: Visualizzazione in Protegè dell'individuo mappato

6 Applicazione Web

In merito allo sviluppo del client Javascript, è stata sviluppata una *single-page application* eseguibile su un web server [NodeJS](#) utilizzando le librerie grafiche [ReactJS](#), [Bootstrap](#) e [Bootstrap Icons](#), le quali ci hanno concesso una maggiore velocità e flessibilità nella fase di sviluppo.

Soluzione proposta L'applicazione essenzialmente implementa un *faceted browser*, fornendo all'utente due principali funzionalità: la ricerca di Claim e la visualizzazione delle Review e delle Mention legate a uno specifico Claim. Al fine di realizzare tali funzionalità, l'applicazione si presenta all'utente articolata nelle due pagine **Home**, **Results** e nei cinque componenti React **Home**, **Results**, **ClaimCard**, **ReviewRichSnippet**, **MentionCard**.

La pagina **Home** corrisponde all'omonimo componente, raggiungibile all'URL `"/"`. Com'è visibile in Figura 19, essa è suddivisa in tre sezioni: la prima, in cui vi è un messaggio di benvenuto; la seconda, in cui è possibile effettuare la ricerca in modo basilare oppure avanzato con più filtri; e infine la terza, in cui è presente una tabella con un ranking dei claimants per numero di Claim. Compilando il form di ricerca e premendo sul bottone "Search" presente al fondo della pagina, l'utente verrà reindirizzato alla pagina **Results**, all'interno della quale verranno caricati i risultati corrispondenti alla richiesta di ricerca.

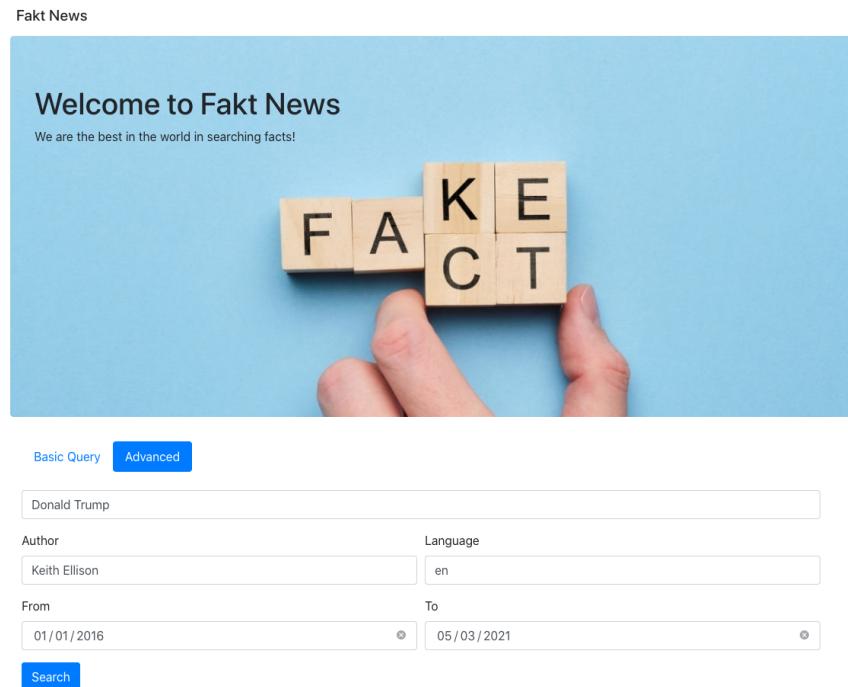


Figura 19: Renderizzazione del componente **Home**, corrispondente alla pagina **Home** reperibile all'URL `"/"`. Non è visibile la tabella con il ranking dei claimant poiché non rientrava nello screenshot della pagina.

La pagina **Results** corrisponde (come per la pagina **Home**) all'omonimo componente, all'interno del quale vengono renderizzati i componenti **ClaimCard**, **ReviewRichSnippet** e **MentionCard**. In questa pagina raggiungibile all'URL `"/search"` e visibile in Figura 20 vengono visualizzati i risultati della ricerca effettuata nella **Home**, divisi in tre colonne. Nella prima viene visualizzata la lista di

Claim risultanti dalla ricerca precedente, renderizzata tramite l'uso di componenti **ClaimCard**, nella seconda colonna vengono visualizzate le Review relative a un Claim selezionato, renderizzate tramite l'uso di componenti **ReviewRichSnippet**, e nella terza vengono visualizzate le Mention delle entità presenti all'interno della Review, renderizzate grazie all'ausilio di componenti **MentionCard**.

The screenshot shows a search results page for 'Fakt News'. It displays three main items, each with a title, date, and a brief description. The first item is from 'Keith Ellison' on '26 Jun 2018' about the US Supreme Court's shutdown of Muslims entry into the United States. The second and third items are from 'FactCheck.org' on '25 Aug 2017' and '09 Mar 2017' respectively, both regarding President Donald Trump's travel ban. Each item has a 'ClaimCard' button. To the right of the third item, there is a detailed view of the FactCheck.org claim, which includes a 'ReviewRichSnippet' (with a green dashed box around it) and a 'MentionCard' (with a blue dashed box around it). The 'ReviewRichSnippet' contains a photograph of wooden blocks spelling 'FACT' and text about the travel ban. The 'MentionCard' contains information about the United States and linked resources.

Figura 20: Renderizzazione del componente Results, corrispondente alla pagina Results reperibile all'URL "/search". Al suo interno è stata evidenziata la presenza dei componenti **ClaimCard**, **ReviewRichSnippet**, **MentionCard** grazie all'uso dei colori rosso, verde e blu.

Scambio dati con la piattaforma LOD Per renderizzare i dati, i componenti hanno bisogno di collegarsi al repository GraphDB `modsem-faktnews`, al quale invieranno le query costruite in **modo programmatico** dal client in base ai campi presenti nei filtri di ricerca. Per soddisfare tale scopo, ai componenti, è sufficiente importare il file `TripleStoreConfig.js`, che contiene la configurazione del repository GraphDB e i prefissi predefiniti da aggiungere in ogni query.

Per interagire con il repository GraphDB, `TripleStoreConfig.js` utilizza la libreria `@innotrade/enapso-graphdb-client` (reperibile sia come Node Module oppure su Github al seguente [link](#)), la quale fornisce delle API per l'interrogazione del repository tramite chiamate HTTP. Questa libreria è responsabile dell'incapsulamento della query in una richiesta POST e dell'invio di quest'ultima al repository GraphDB tramite HTTP. Una volta effettuata la computazione, il repository inoltrerà tramite una risposta HTTP il result set della query alla libreria, la quale incapsulerà tale result set in un JSON per consentire ai componenti di renderizzare i dati restituiti. Nel listato 1 è presente un esempio di formattazione del file JSON in caso di query eseguita con successo, mentre nel listato 2 è presente un esempio di formattazione del JSON in caso di errore restituito dal repository.

Query SPARQL implementate È opportuno notare che nello sviluppo dell'applicazione non sono state implementate tutte le query SPARQL descritte nella sezione 4.1, piuttosto, sono state usate solamente le seguenti query:

1. Q4 - "Contextual Info", query federata a DBpedia,

```
{
  "total": 3,
  "success": true,
  "records": [
    {
      "claim_author": "FactCheck.org",
      "claim": "C-48",
      "language": "en",
      "claim_date": "2018-01-04",
      "content": "President Barack Obama repeatedly [...]"
    },
    {
      "claim_author": "FactCheck.org",
      "claim": "C-59",
      "language": "en",
      "claim_date": "2018-01-19",
      "content": "Denzel Washington said Barack Obama was a [...]"
    },
    {
      "claim_author": "Charles Boustany",
      "claim": "C-43",
      "language": "en",
      "claim_date": "2016-08-25",
      "content": "Barack Obama and Hillary Clinton refuse [...]"
    }
  ]
}
```

Listing 1: JSON elaborato dalla libreria a partire dal result set della query di ricerca. In questo caso, l'utente ha cercato tutti i Claim in cui compariva la parola chiave "Barack Obama". Il contenuto dei Claim è stato troncato per favorire la leggibilità.

```
{
  "success": false,
  "message": "401 - Bad credentials",
  "statusCode": 401
}
```

Listing 2: JSON elaborato dalla libreria in seguito a un'errore. È importante osservare che la libreria codifica anche il messaggio di errore restituito dal server, informazione preziosa in fase di debug.

2. Q6 - "Check Trusted Author", leggermente modificata per venire incontro alle esigenze implementative del client Javascript,
3. Q7 - "Get related Reviews",
4. Q10 - "Claimants Ranking",
5. Una versione modificata di Q1 ("Review search by keyword") fusa con Q3 ("Claim Temporal Filter"), visibile nel metodo `handleSearchQuery()` del componente Home. La fusione delle due query ha permesso di effettuare una ricerca sui Claim usando dei filtri avanzati come l'autore della review, il linguaggio e l'intervallo di tempo dei Claim. Come per Q6, la modifica è stata effettuata solamente con l'obiettivo di venire incontro alle esigenze implementative del client.

7 Conclusioni

Nella sezione 3 sono state presentate alcune assunzioni fatte alla base del modello. Di conseguenza alcune delle semplificazioni effettuate, potrebbero essere sviluppate in versioni successive, per aumentare l'espressività dell'ontologia e il suo pratico utilizzo. dal punto di vista dell'applicazione web proposta, in futuro sarebbe possibile arricchire quest'ultima, aggiungendo una pagina di informazioni riguardo al claimant di un dato Claim, all'interno della quale si potrebbero inserire le query Q5 ("Get claimant credibility") e Q8 ("Get last Claims"), oppure ancora si potrebbero implementare le query Q9 ("Social Account") all'interno della MentionCard (con opportune modifice di UI) e la Q2 ("Unreviewed Claims") in una sezione della pagina Home.

Grazie allo sviluppo end-to-end iniziato dal modello concettuale e concluso nello sviluppo dell'applicazione web di Fakt-News, abbiamo potuto apprezzare quanto e come le tecnologie basate sul Semantic Web, possano essere di supporto alla realizzazione di sistemi informativi di nuova generazione.

8 Appendice

Introducing IMVAIN

For the purposes of this course, we use the acronym IMVA/IN to methodically evaluate sources who show up in news stories.

I:

Independent sources are better than self-interested sources

M:

Multiple sources are better than single sources

V:

Sources who **Verify** with evidence are better than sources who assert

A / I:

Authoritative / Informed sources are better than uninformed sources

N:

Named sources are better than unnamed sources

Figura 21: Principi alla base del Metodo IMVAIN

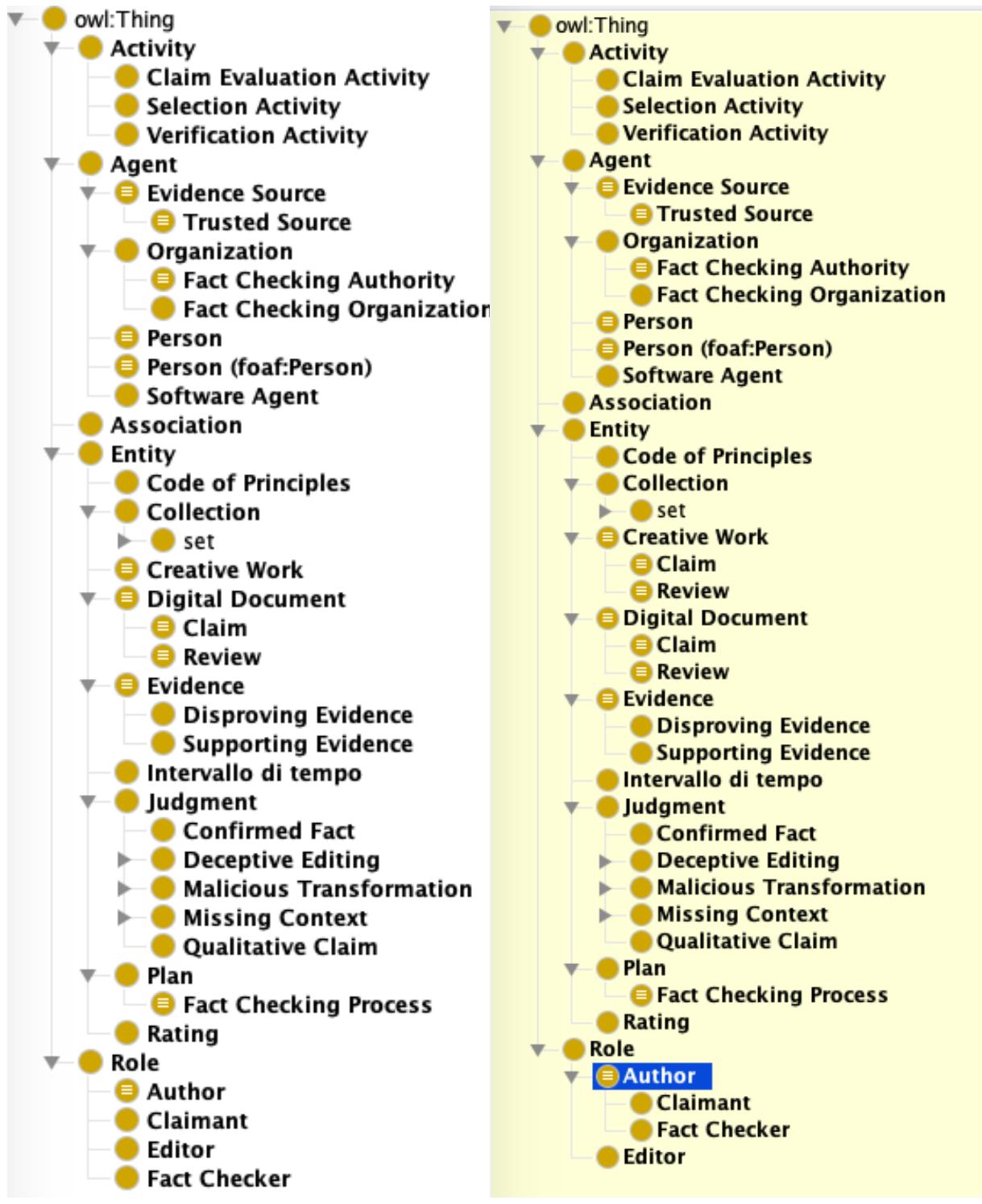


Figura 22: Tassonomia delle classi.

2 Analyse author / source

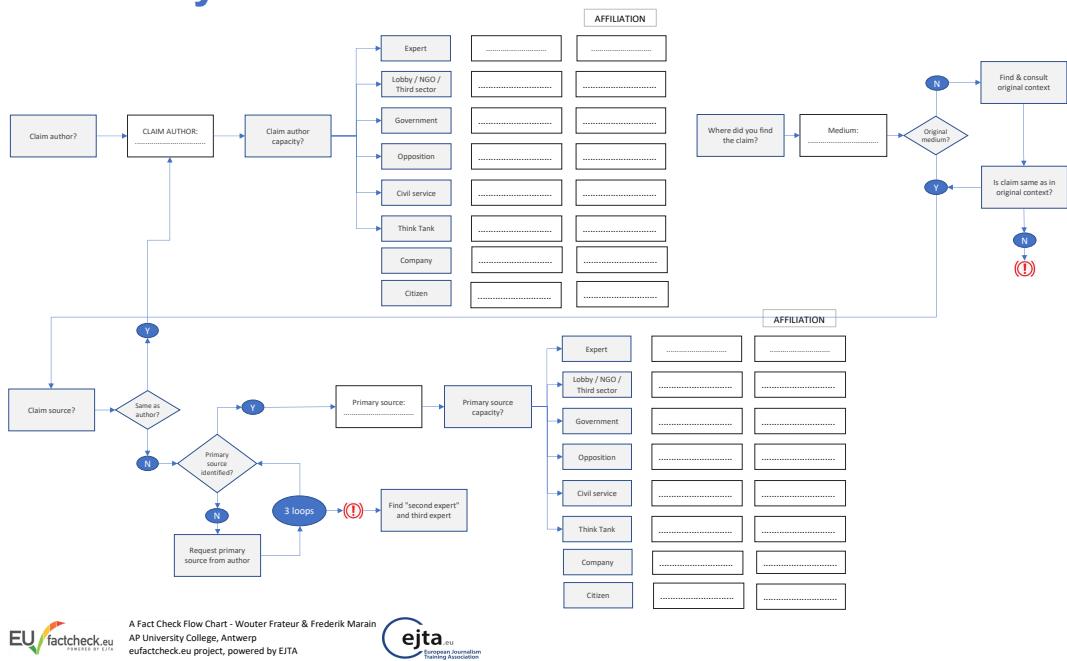


Figura 23: Flowchart per che guida la scelta di una sorgente dell'evidenza. Questo flowchart fa parte di una più ampia collezione di materiale educativo disponibile su eufactcheck.eu

```

<urn:MAPID-Claim_1> a rr:TriplesMap;
  rr:logicalTable [ a rr:R2RMLView;
    rr:sqlQuery """SELECT claim.id AS cid,
      claim.content AS ccn,
      claim.publication_date AS cpd,
      claim.url AS curl,
      claim.language AS cla,
      review.id AS rid
    FROM claim, about, review
    WHERE review.id = about.review_id AND about.claim_id = claim.id ;"""
  ];
  rr:predicateObjectMap [ a rr:PredicateObjectMap;
    rr:objectMap [ a rr:ObjectMap, rr:TermMap;
      rr:template "http://www.modsem.org/fakt-news#R-{rid}";
      rr:termType rr:IRI
    ];
    rr:predicate :associatedReview
  ], [ a rr:PredicateObjectMap;
    rr:objectMap [ a rr:ObjectMap, rr:TermMap;
      rr:column "ccn";
      rr:termType rr:Literal
    ];
    rr:predicate :claimContent
  ], [ a rr:PredicateObjectMap;
    rr:objectMap [ a rr:ObjectMap, rr:TermMap;
      rr:column "cpd";
      rr:termType rr:Literal
    ];
    rr:predicate dct:date
  ], [ a rr:PredicateObjectMap;
    rr:objectMap [ a rr:ObjectMap, rr:TermMap;
      rr:column "curl";
      rr:datatype xsd:anyURI;
      rr:termType rr:Literal
    ];
    rr:predicate schema:url
  ], [ a rr:PredicateObjectMap;
    rr:objectMap [ a rr:ObjectMap, rr:TermMap;
      rr:column "cla";
      rr:termType rr:Literal
    ];
    rr:predicate dct:language
  ];
  rr:subjectMap [ a rr:SubjectMap, rr:TermMap;
    rr:class :Claim;
    rr:template "http://www.modsem.org/fakt-news#C-{cid}";
    rr:termType rr:IRI
  ] .

```

Listing 3: Specifica R2RML per il mapping MAPID-Claim.

Riferimenti bibliografici

- [1] Michael Glanzberg. «Truth». In: *The Stanford Encyclopedia of Philosophy*. A cura di Edward N. Zalta. Fall 2018. Metaphysics Research Lab, Stanford University, 2018.
- [2] “Fake news” is 2017 American Dialect Society word of the year. en-US. Gen. 2018. URL: <https://www.americandialect.org/fake-news-is-2017-american-dialect-society-word-of-the-year> (visitato il 12/12/2020).
- [3] *Collins 2017 Word of the Year Shortlist*. en-GB. Nov. 2017. URL: <http://blog.collinsdictionary.com/language-lovers/collins-2017-word-of-the-year-shortlist/> (visitato il 12/12/2020).
- [4] *Macquarie Dictionary*. URL: <https://www.macquariedictionary.com.au/resources/view/word/of/the/year/2016> (visitato il 12/12/2020).
- [5] Edson C. Tandoc, Zheng Wei Lim e Richard Ling. «Defining “Fake News”: A typology of scholarly definitions». en. In: *Digital Journalism* 6.2 (feb. 2018), pp. 137–153. ISSN: 2167-0811, 2167-082X. DOI: <10.1080/21670811.2017.1360143>. URL: <https://www.tandfonline.com/doi/full/10.1080/21670811.2017.1360143> (visitato il 12/12/2020).
- [6] Jackie Mansky. *The Age-Old Problem of “Fake News”*. en. URL: <https://www.smithsonianmag.com/history/age-old-problem-fake-news-180968945/> (visitato il 12/12/2020).
- [7] Cherilyn Ireton, Julie Posetti e UNESCO. *Journalism, "fake news" et disinformation: handbook for journalism education and training*. English. OCLC: 1053833113. 2018. ISBN: 9789231002816. URL: <http://unesdoc.unesco.org/images/0026/002655/265552E.pdf> (visitato il 06/12/2020).
- [8] Bill Kovach e Tom Rosenstiel. *The elements of journalism: what newsmen should know and the public should expect*. Revised and updated third edition. New York: Three Rivers Press, 2014. ISBN: 978-0-8041-3678-5.
- [9] Poynter Institute. *IFCN Code of Principles*. URL: <https://www.ifcncodeofprinciples.poynter.org/know-more/the-commitments-of-the-code-of-principles> (visitato il 09/01/2021).
- [10] Andon Tchetchmedjiev et al. «ClaimsKG: A Knowledge Graph of Fact-Checked Claims». In: *The Semantic Web – ISWC 2019*. A cura di Chiara Ghidini et al. Cham: Springer International Publishing, 2019, pp. 309–324. ISBN: 978-3-030-30796-7.
- [11] *The ClaimReview Project*. en-US. URL: <https://www.claimreviewproject.com> (visitato il 09/01/2021).
- [12] *The methodologies of fact-checking*. en. URL: https://ballotpedia.org/The_methodologies_of_fact-checking (visitato il 06/03/2021).
- [13] *The State of Automated Factchecking*. en. Ago. 2016. URL: <https://fullfact.org/blog/2016/aug/automated-factchecking/> (visitato il 06/03/2021).
- [14] Sunanda Creagh. *The Conversation’s FactCheck granted accreditation by International Fact-Checking Network at Poynter*. en. URL: <http://theconversation.com/the-conversations-factcheck-granted-accreditation-by-international-fact-checking-network-at-poynter-74363> (visitato il 06/03/2021).
- [15] Sylvie Cazalens et al. «A Content Management Perspective on Fact-Checking». In: *Companion Proceedings of the The Web Conference 2018*. WWW ’18. Lyon, France: International World Wide Web Conferences Steering Committee, apr. 2018, pp. 565–574. ISBN: 9781450356404. DOI: <10.1145/3184558.3188727>. URL: <https://doi.org/10.1145/3184558.3188727> (visitato il 06/03/2021).
- [16] Michelle A. Amazeen. «Revisiting the Epistemology of Fact-Checking». In: *Critical Review* 27.1 (gen. 2015), pp. 1–22. ISSN: 0891-3811. DOI: <10.1080/08913811.2014.993890>. URL: <https://doi.org/10.1080/08913811.2014.993890> (visitato il 06/03/2021).
- [17] Tariq Alhindi, Savvas Petridis e Smaranda Muresan. «Where is Your Evidence: Improving Fact-checking by Justification Modeling». In: *Proceedings of the First Workshop on Fact Ex-*

- traction and VERification (FEVER)*). Brussels, Belgium: Association for Computational Linguistics, nov. 2018, pp. 85–90. DOI: [10.18653/v1/W18-5513](https://doi.org/10.18653/v1/W18-5513). URL: <https://www.aclweb.org/anthology/W18-5513>.
- [18] Isabelle Augenstein et al. «MultiFC: A Real-World Multi-Domain Dataset for Evidence-Based Fact Checking of Claims». In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, nov. 2019, pp. 4685–4697. DOI: [10.18653/v1/D19-1475](https://doi.org/10.18653/v1/D19-1475). URL: <https://www.aclweb.org/anthology/D19-1475>.