



Metodologie, strumenti e servizi innovativi per lo sviluppo del patrimonio culturale dei Geo-Archeo-Siti

OR4 - Fruizione

Attività 4.1 - Servizi integrati per il servizio geo-archeologico,
strumenti e modelli per la fruizione dei G.A.S.

**REPORT - Individuazione di modelli e sistemi per la gestione
del profilo degli utenti costruzione**

TITOLO: Individuazione di modelli e sistemi per la gestione del profilo degli utenti

OBIETTIVO REALIZZATIVO: OR1 Fruizione

DATA DI SCADENZA: 31/05/2022

DATA DI PRESENTAZIONE: 31/05/2022

PARTNER RESPONSABILE: T&G Srl

PARTNER CONTRIBUTORI: DING – Università del Sannio

STATO: Finale

AUTORI: Vari

REVISORI: Maria Tortorella

NATURA: Report

LIVELLO DI DIFFUSIONE: Pubblico

CRONOLOGIA DELLE VERSIONI

Versione	Data	Autore	Partner	Descrizione
1.0	15/04/2022	Vari	T&G	Prima stesura
1.1	10/05/2022	Vari	T&G	Seconda stesura
2.0	31/05/2022	Vari	T&G	Versione definitiva

Individuazione di modelli e sistemi per la gestione del profilo degli utenti

INDICE

Preambolo

1 Modellazione, costruzione e gestione di Profili utente

1.1 I contributi provenienti dalla letteratura

1.1.1 Modellazione del profilo utente

1.1.2 Costruzione del profilo utente

1.1.3 Gestione del profilo utente

1.2 Analisi critica

1.3 Gestione dei profili utente, cloud e big data

1.3.1 Cloud e Big Data

1.3.2 Servizio di gestione dei profili utente

Bibliografia di riferimento

2 Approcci alla profilazione degli utenti

2.1 Schema di classificazione della profilazione degli utenti

2.2 Raccolta dei dati

2.2.1 Contenuto del profilo utente

2.2.2 Approcci alla raccolta dei dati

2.3 Costruzione e modellazione del profilo utente

2.3.1 Rappresentazione del profilo utente

2.3.2 Costruzione del profilo utente

2.4 Personalizzazione

2.4.1 Identificazione dell'utente

2.4.2 Applicazioni

2.5 Analisi critica

Bibliografia di riferimento

Preambolo

I modelli e i sistemi per la gestione del profilo degli utenti sono utilizzati dalle piattaforme digitali per raccogliere, archiviare e utilizzare le informazioni degli utenti registrati. Questi modelli e sistemi possono essere utilizzati per una vasta gamma di applicazioni, come la personalizzazione dei contenuti, la profilazione comportamentale, la pubblicità mirata, la gestione delle preferenze e delle autorizzazioni degli utenti.

Uno dei modelli più comuni per la gestione del profilo degli utenti è il modello dati utente, che prevede la creazione di un database per ogni utente registrato, in cui sono memorizzate le informazioni relative al loro profilo. Questo database può contenere informazioni personali dell'utente, come nome, cognome, indirizzo, e-mail, numero di telefono, ma anche informazioni sulle preferenze, gli interessi, le attività e le interazioni dell'utente sulla piattaforma.

Altri modelli di gestione del profilo degli utenti possono includere tecniche di profilazione comportamentale, che utilizzano algoritmi di machine learning per analizzare il comportamento degli utenti e creare profili in base ai loro interessi, alle loro abitudini di navigazione e alle loro interazioni sulla piattaforma.

Per quanto riguarda i sistemi di gestione del profilo degli utenti, questi possono essere implementati attraverso l'uso di software specializzati o attraverso la creazione di sistemi personalizzati in-house. Alcune piattaforme digitali possono utilizzare anche API di terze parti per raccogliere e gestire i dati degli utenti. È importante sottolineare che la gestione del profilo degli utenti deve essere sempre effettuata nel rispetto delle normative sulla privacy e della sicurezza dei dati degli utenti. Le piattaforme digitali devono adottare misure di sicurezza adeguate a proteggere i dati degli utenti da accessi non autorizzati o da violazioni della sicurezza dei dati. Inoltre, le piattaforme digitali devono garantire ai propri utenti la possibilità di controllare le informazioni raccolte su di loro e di scegliere come queste informazioni vengono utilizzate.

3 Modellazione, costruzione e gestione di Profili utente

Il miglioramento dell'accuratezza dei risultati delle applicazioni e dei sistemi informatici e la soddisfazione delle esigenze degli utenti in termini di conoscenza è sempre stato un compito molto complicato e rappresenta una delle sfide più grandi per i ricercatori, perché la **conoscenza** deve essere estratta, raccolta, filtrata, convalidata, gestita, prevista e normalizzata per essere rappresentata in modo tale da essere sfruttata e, naturalmente, deve essere protetta.

I diversi formati dei dati, la mancanza di interoperabilità, la difficoltà di garantire l'interoperabilità semantica e la distribuzione geografica dei dati, oltre all'emergere di nuovi paradigmi informatici come i Big Data e il cloud computing, rendono la conoscenza molto complicata da acquisire.

Un profilo utente è una **conoscenza** relativa a un utente specifico (contesto, interessi, preferenze, ecc.), un vincolo decisivo preso in considerazione e sfruttato dai ricercatori per orientare meglio i risultati delle applicazioni, che consentono di soddisfare le aspettative degli utenti e, in alcuni casi, di suggerire raccomandazioni pertinenti.

La profilazione degli utenti è un'area molto interessante, situata all'intersezione tra Intelligenza Artificiale, Data mining e metodi di apprendimento e che ha impegnato severamente molti ricercatori.

Al giorno d'oggi, la profilazione degli utenti è vitale in molti settori in cui è essenziale conoscere gli utenti di applicazioni software (Schiaffino et al, 2009), quali: e-recruitment, e-commerce, criminalità, sistemi di

raccomandazione adattivi, information retrieval, istruzione, sicurezza, e-learning, (Soniya et al, 2013; kaczmarek et al, 2010; Alguliev et al, 2014; Jeong et Choi, 2012; Satler et al, 2010), come notato in (Kanoje et al, 2014): sistemi di personalizzazione (Bedekar et al, 2008), (Doan et al, 2002), siti web adattivi, negozi web adattivi, siti web di base per il turismo elettronico (Ouanaim et al, 2010), e-recruitment (Bradley et al, 2000), sistemi decisionali (Aquino e Filgueiras, 20) e molti altri.

Gli utenti interagiscono ogni giorno con molte nuove applicazioni, che girano su piattaforme diverse in ambienti diversi, richiedono informazioni diverse, hanno obiettivi diversi e forniscono servizi diversi. Negli ultimi anni sono stati compiuti molti sforzi per scoprire e rappresentare le caratteristiche di base degli utenti (Salter et al, 2010).

Nel Web, nell'Ubiquitous e anche nel Cloud computing, sono stati proposti molti sistemi per raccogliere, organizzare e gestire le informazioni degli utenti analizzando le loro diverse interazioni con varie applicazioni e servizi Web, sfruttando diverse tecnologie di modellazione e metodi di gestione della conoscenza.

Sono stati fatti molti lavori sulla profilazione degli utenti e di seguito è presentata una rassegna completa ed esaustiva dei più importanti approcci e meccanismi sviluppati per la profilazione degli utenti.

Per fare ciò, in una prima parte, si è suddiviso il processo di profilazione in tre fasi principali:

1. la modellazione,
2. la costruzione e
3. la gestione dei profili utente,

discutendo le diverse tecniche e i metodi utilizzati per ciascuna fase.

In una seconda parte, sono presentati i nuovi paradigmi IT: Cloud e Big Data, e proposta un'architettura per la gestione dei profili utente sul Cloud computing.

La presente Sezione è organizzata come segue:

La Parte 1 presenta i contributi provenienti dalla letteratura suddivisi in tre aree: modellazione, costruzione e gestione dei profili utente nei diversi ambienti (Web, Ubiquitous, Cloud), mentre nella Parte 2 viene presentata una analisi critica dei diversi metodi. La Parte 3 presenterà le nozioni più importanti sul Cloud computing e sui Big Data. Poi, sono illustrati i diversi moduli per una architettura "ideale" per la gestione dei profili utente.

1.4 I contributi provenienti dalla letteratura

La tabella 1 mostra la classificazione adottata per presentare una panoramica dei principali lavori svolti nell'ambito della profilazione degli utenti.

In questa Parte il processo di profilazione è suddiviso in tre fasi principali: modellazione del profilo utente, costruzione del profilo utente e gestione del profilo utente, come segue:

gestione del profilo utente	modellazione del profilo utente
	costruzione del profilo utente

Tabella 1

1.4.1 Modellazione del profilo utente

Negli ultimi anni sono stati fatti molti sforzi e diversi lavori sulla rappresentazione delle caratteristiche di base dell'utente, e sono stati proposti molti approcci volti a modellare le informazioni dell'utente al fine di rispondere ai suoi bisogni, esigenze e desideri (Needs, Requirements and Desires (NRD) (Salem e Rauterberg, 2004) su un dato sistema (Liu et al, 2009).

Il profilo dell'utente viene utilizzato per eliminare l'ambiguità, guidare le azioni del sistema e migliorare la qualità delle interazioni con l'utente (Kay, 2001).

È possibile dividere le informazioni dell'utente in due categorie: informazioni statiche e dinamiche (Iggui et al, 2016), che variano da un dominio applicativo all'altro, da semplici informazioni demografiche personali a informazioni complesse come le conoscenze di base, gli interessi, le preferenze, gli obiettivi, i comportamenti, le caratteristiche individuali dell'utente e il contesto dell'utente, ecc (Schiaffino et al, 2009) (Liu et al, 2009). Questo contenuto dipende dal campo di applicazione, che è fortemente legato ai meccanismi e agli strumenti utilizzati per la modellazione.

In letteratura si distinguono diverse tecniche per la modellazione dell'utente che permettono di risolvere molte sfide indifferenti agli ambienti (web, ubiquo, cloud), come l'eterogeneità, l'interoperabilità semantica, la scalabilità e l'espressività del supporto, ecc.

Di seguito si illustrano i più utilizzati dai ricercatori recenti: (Kuflik et al, 2012) (Schmohl e Baumgarten, 2008) (Strang et al, 2004) (Vanathi et al, 2011):

- ~ *Modelli basati sulle parole chiave*: noti come *Bag-of-keywords* e rappresentazione vettoriale, il modello più semplice e comune, considerato come una rappresentazione di base, in cui l'utente è rappresentato da un insieme di termini o concetti, eventualmente ponderati (Kuflik et al, 2012) (Malone et al, 1987). In genere, le parole chiave vengono estratte da documenti e fonti testuali consultate e ponderate utilizzando TF/IDF (Manning et al, 2008). I concetti rappresentano gli interessi dell'utente e i pesi rappresentano il grado di interesse. I valori possono anche essere binari (0 o 1), per indicare comportamenti come l'acquisto o meno, il clic o meno, oppure possono essere numeri interi (Hafez e Xu, 2013). In (Alguliev et al, 2014), gli autori hanno lavorato sul rilevamento di un *masquerader* nell'ambiente cloud. Per scoprire un attacco illegale, viene creato un profilo di comportamento per ogni utente sotto forma di vettore che tiene conto delle caratteristiche dell'utente e dei suoi interessi; il comportamento normale di un utente viene modellato come un modello cloud, quindi gli altri modelli vengono confrontati e vengono valutate le deviazioni da questo comportamento; se un valore di deviazione è superiore alla soglia, l'utente che cerca di accedere al sistema viene valutato come utente illegale. In (Tayari et al, 2009) gli autori hanno proposto un modello multidimensionale basato su una rappresentazione algebrica delle emozioni degli utenti in un cosiddetto "Spazio Victor", dove le emozioni sono rappresentate da vettori.

Questo tipo di rappresentazioni classiche è molto utilizzato nell'Information Retrieval (IR), in particolare per la modellazione di informazioni statiche e dinamiche sull'utente provenienti da fonti esplicite e implicite. (Liu et al, 2009).

- ~ *Modelli Markup scheme*: Consistono in strutture di dati gerarchiche basate su tag di markup, compresi attributi e commenti. Di solito sono implementati come derivati di SGML1 come XML - eXtensible Markup Languages.

In alcuni casi, i modelli di schemi di markup sono utilizzati per descrivere il contesto come estensioni del Composite Capabilities / Preferences Profile (CC/PP)² e dell'User Agent Profile (UAProf)³ per coprire l'elevata dinamica delle informazioni contestuali (Schmohl & Baumgarten, 2008).

Per garantire l'interoperabilità e la portabilità, diversi lavori di modellazione del profilo dell'utente si

¹ Standard Generalized Markup Language (iso 8879:1986). <http://www.iso.org>

² Composite Capabilities/ Preferences Profile (CC/PP). <http://www.w3.org/Mobile/CCPP/>

³ User Agent Profile. <http://www.wapforum.org>

basano su linguaggi strutturati, come (Abel et al, 2011) che, basandosi sul grafo RDF - Resource Description Framework che collega i post di Twitter con gli articoli, hanno rappresentato l'utente come una combinazione di due profili: Profilo basato su argomenti che include gli interessi dell'utente (sport, politica, ecc.) e Profilo basato su entità che rappresenta gli interessi dell'utente come persone, organizzazioni o eventi. (EARL, 2006) basato su XML per la modellazione delle emozioni degli utenti. (Jeong e Choi, 2012) in cui gli autori hanno proposto un modello basato su XML, che tiene conto delle informazioni sull'utente, sui servizi e sul dispositivo, per l'autenticazione dell'utente nel cloud computing.

- ~ *Approcci basati sulle ontologie:* Un'ontologia è una specifica di una concettualizzazione (Gruber, 2008) e fornisce quindi un mezzo per l'organizzazione e la gestione della conoscenza in modo tale da essere condivisibile e riutilizzabile tra le applicazioni di diversi domini. Una volta che la conoscenza è organizzata sotto forma di ontologia, la capacità di inferenza fornisce la possibilità di estrarre relazioni implicite dagli interessi degli utenti. L'ontologia è uno strumento potente per rappresentare e condividere la conoscenza sugli utenti, sui gruppi di utenti o anche su un dominio (Han e Chang, 2002) (King et al, 2007).

Per questo, in molti lavori i ricercatori utilizzano diverse ontologie di riferimento come l'ontologia FOAF, il Wordnet Dictionary, l'ODP (Daoud e Tamine, 2008), Open Directory e persino Wikipedia per categorizzare i concetti, il che facilita l'interazione semantica di un gran numero di applicazioni diverse con la stessa base di conoscenza (Wasim et al, 2011).

Gli interessi e la conoscenza di base dell'utente sono rappresentati da concetti o argomenti, come in (Michelson et al, 2010), che utilizza Wikipedia per disambiguare le entità evocate dai tweet dell'utente e per mapparle alle categorie che definiscono il profilo dell'argomento, e (Umamaheswariand Patil, 2012), in cui la conoscenza (interessi) è estratta da documenti archiviati, pagine web sfogliate ed e-mail composte/ricevute. Indicate come il modello più espressivo (Strang et al, 2004) (Couto et al, 2005), le ontologie sono utilizzate in molti campi come la Web Intelligence, i sistemi di raccomandazione e, naturalmente, l'Information Retrieval.

- ~ *Modelli basati su grafici:* Tali modelli consistono nel rappresentare graficamente le entità e le loro relazioni; a questo proposito sono stati sviluppati molti linguaggi di modellazione, come l'Unified Modeling Language UML⁴ e l'Object-Role Modeling ORM⁵, che rappresentano l'utente e il suo contesto identificando i fatti e arricchendoli con tipi e ruoli (Schmohl & Baumgarten, 2008). Nell'ambito del progetto APMD⁶ (Personalization Access to Mass of Data) (Bouzeghoub e Kostadinov, 2005) (Kostadinov, 2005) è stato discusso e proposto un approccio di modellazione del profilo dell'utente per la gestione della conoscenza e per adattare l'esecuzione delle query al fine di ottenere migliori risultati di personalizzazione. Gli autori esprimono le preferenze dell'utente come una combinazione di informazioni sull'utente (interessi, dati personali, dati di sicurezza) e di informazioni sul contesto in un generico ed estensibile modello multidimensionale che fornisce un supporto per una piattaforma di gestione dei profili, consentendo l'importazione/esportazione di profili utente, la specializzazione/istituzione di profili, la corrispondenza e l'evoluzione dei profili utente.

Per gestire l'ambiguità generata da diverse fonti di contesto dell'utente, (Sklab et al, 2013), ispirandosi a (Baldauf et al, 2007) (Chaari, 2007) (Dey et al, 1999), hanno proposto un modello generico di contesto dell'utente, istanziabile in diversi domini in ambienti pervasivi (e-business,

⁴ Unified Modeling Language, <http://www.uml.org>

⁵ Object Role Modeling, <http://www.orn.net>

⁶ ACI Masses de Données : Projet MD-33 Accès Personnalise à des Masses de Données. <http://apmd.prism.uvsq.fr/>

campo medico) utilizzando un formalismo grafico, ontologie logiche per la descrizione asemantica e XML per la memorizzazione. Troviamo anche (Kankainen e Parkinen, 2001) che hanno proposto una rappresentazione grafica del profilo dell'utente (GUP) per i servizi di telecomunicazione, utilizzando la stessa idea di (Jonassen et al, 1993), dove gli autori tengono conto del background dell'utente, del suo contesto, del suo settore di vita e delle immagini.

- ~ *Modelli basati sul Machine Learning*: Il modello dell'utente viene sviluppato sulla base di dati di addestramento, invisibili per gli utenti. Questo approccio non è molto utilizzato perché richiede un'ampia quantità di dati per la modellazione degli interessi degli utenti. Inoltre, la natura dinamica degli interessi degli utenti richiede un processo di apprendimento continuo, cosa che di solito non avviene nei metodi di apprendimento automatico degli interessi degli utenti (Michelson et al, 2010).
- ~ *Modelli basati sulla logica*: rappresentano approcci di modellazione altamente formali, basati su logiche che definiscono le condizioni in base alle quali le espressioni conclusive sui fatti possono essere derivate da insiemi di altre espressioni o fatti. Queste condizioni sono descritte da regole in un sistema formale, in modo che i fatti, le espressioni e le regole insieme definiscano il contesto (Schmohl e Baumgarten, 2008).
- ~ *Modelli Object-oriented*: Consistono nell'incapsulare le informazioni contestuali in oggetti. (Schmohl e Baumgarten, 2008).

1.4.2 Costruzione del profilo utente

La profilazione degli utenti è ormai presente nella maggior parte dei sistemi informativi; la motivazione della creazione di profili utente è che gli utenti differiscono per preferenze, interessi, background e obiettivi nell'uso delle applicazioni software.

Scoprire queste differenze è fondamentale per fornire agli utenti servizi personalizzati (Schiaffino et al, 2009). La costruzione di un profilo utente dipende dalle fonti di informazione e dal campo di applicazione. (Kanoje et al, 2014) hanno presentato un'indagine sulle tendenze e sulle tecniche di profilazione degli utenti; gli autori hanno identificato tre tipi di approccio alla profilazione degli utenti in base all'implicazione dell'utente nel processo di profilazione.

- a) *Approcci espliciti*: si basano sulle informazioni fornite dall'utente, di solito attraverso moduli di registrazione sul Web, questionari o strumenti psicometrici appositamente progettati. Ma rivelare le proprie informazioni a ogni sistema informativo è diventato noioso. La stragrande maggioranza degli utenti è riluttante a fornire informazioni esplicite sulle proprie preferenze e aree di interesse (Salter et al, 2010). Per questo i ricercatori hanno presentato
- b) *Approcci impliciti*: basandosi sui comportamenti degli utenti e tenendo conto del vincolo della dinamicità, in letteratura sono state proposte molte metodologie, tecniche e metodi di apprendimento per prevedere le informazioni sull'utente e sulle sue esigenze, come spiegato da (Poo et al, 2003).
- c) *Approcci ibridi*: questo tipo combina i vantaggi di entrambi gli approcci precedenti.

Di seguito si è maggiormente focalizzati sugli approcci impliciti in cui il profilo dell'utente è costruito da un sistema di profilazione non supervisionato. Secondo (Iggui et al, 2016), (Kanoje et al, 2014) e (Jie et al, 2010) il processo di profilazione degli utenti è articolato in quattro fasi:

1) *Raccolta dei dati*: attraverso l'estrazione di informazioni sull'utente da fonti diverse, come l'estrazione di dati dal web (Jie et al, 2010), dai social media (Li et al, 2014), dall'analisi dei comportamenti dell'utente. Ad esempio (Hijikata, 2004) ha tracciato le operazioni del mouse dell'utente per estrarre dal testo delle pagine web le parti di testo a cui l'utente potrebbe essere interessato.

Alcune delle operazioni principali sono: tracciamento del testo, puntamento dei link, selezione del testo,

scorrimento, salvataggio, stampa, movimento e ridimensionamento della finestra. Molti lavori hanno utilizzato misure TF/IDF, ontologie o dizionari per identificare informazioni rilevanti come parole chiave, entità denominate, ecc.

2) Elaborazione dei dati: dopo l'estrazione delle informazioni rilevanti, alcuni dati raccolti possono essere duplicati o incompleti, quindi sono state proposte alcune tecniche per integrare i profili degli utenti e risolvere il problema della disambiguazione, come Hidden Markov Random Field (Jie et al, 2010), machine learning, ontologie e tecniche NLP (Natural Language Processing) (Iggui et al, 2016).

3) Alcuni lavori (Poo et al, 2003), (Stuart et al, 2003) hanno classificato gli utenti in gruppi in base ai loro interessi per guidare meglio il processo di personalizzazione. Questa fase è considerata più come un compito di gestione e sarà discussa nella sezione Gestione dei profili utente.

4) Aggiornamento dei profili: questa fase si occupa dell'evoluzione del profilo dell'utente. Al fine di mantenere un profilo utente fresco e qualitativo (Shmueli-Scheuer et al, 2010). Questa fase ha quindi lo scopo di aggiornare le informazioni dell'utente (Iggui et al, 2016).

La costruzione dei profili degli utenti si colloca all'intersezione di molti campi diversi, utilizzando metodi differenti, dai più semplici ai più complessi, a seconda del tipo di dati e di fonti di informazione: testuali, video, immagini e comportamenti, ecc.

La maggior parte delle informazioni sugli utenti viene estratta dal formato testuale in cui l'utente si esprime o mostra il suo interesse. In (Shmueli-Scheuer et al, 2010), gli autori hanno proposto un framework scalabile in cui i profili degli utenti sono rappresentati dal contenuto testuale utilizzando vettori di *Bag of Words*. Gli autori hanno utilizzato la divergenza di Kullback-Leibler (KL) (Kullback et Leibler, 1951) sui contenuti dei blog degli utenti e hanno implementato il loro approccio su Apache Hadoop MapReduce, ottenendo risultati interessanti, ma molte informazioni vengono perse utilizzando le parole chiave e l'approccio non tiene conto del vincolo semantico.

Per tenere conto di questo aspetto, molti lavori hanno sfruttato la potenza delle ontologie, come ad esempio (Salter et al, 2010) nell'ambito del progetto AGORA dove gli autori hanno utilizzato un insieme di misure statistiche combinate con l'algoritmo presentato da (Widyantoro e Yen, 2001), consentendo la rappresentazione delle preferenze degli utenti con un insieme di concetti ontologici.

(Lau et al, 2009) costruiscono un'ontologia di dominio fuzzy a partire da concetti testuali (blog, e-mail e così via).

(Han e Chen, 2009) hanno proposto un metodo di clustering fuzzy chiamato FCOU che consente all'utente di essere rappresentato da una o più ontologie, permettendo di avere profili arricchiti che saranno più utili al processo di personalizzazione, come in (Umamaheswari e Patil, 2012), dove l'ontologia è costruita estraendo la conoscenza globale dal sistema LCSH (The Library of Congress Subject Headings) e la conoscenza di base dell'utente dai depositi di istanze locali (LIR), come i suoi documenti archiviati, le pagine web sfogliate e le e-mail composte/ricevute. (Sieg et al, 2007) migliorano la qualità dei profili utente assegnando punteggi all'ontologia dei concetti estratti dai documenti consultati in base ai comportamenti dell'utente.

In effetti, i comportamenti degli utenti sono molto indicativi per la costruzione dei loro profili. Diversi lavori si sono occupati dei comportamenti degli utenti, studiandoli e analizzandoli per ottenere maggiori informazioni su di essi. Secondo (Morita e Shinoda, 1994) gli utenti di software passano più tempo a leggere gli articoli a cui sono interessati, mentre gli autori di (Oard e Kim, 1998) studiano altri tipi di comportamenti più rilevanti, come la stampa, il salvataggio, lo scorrimento e il bookmarking, ritenendo che se un utente ha

questi comportamenti durante la navigazione di un documento, deve avere degli interessi in esso. Questa idea è stata adottata da (Xing et al, 2011) nel proporre (BUIE) un algoritmo per l'estrazione dell'interesse dell'utente basato sul suo comportamento. I comportamenti dell'utente possono essere osservati anche durante la sua interazione con i servizi multimediali (Cloud), per ottenere il contenuto di suo interesse invece di cercarlo con la navigazione. Gli autori in (Stupar et al, 2014) utilizzano le informazioni DTV memorizzate nel database del fornitore di contenuti per costruire una cronologia di visione (evento, durata della visione) per generare un profilo utente in grado di personalizzare il fornitore di contenuti. L'approccio potrebbe essere più efficiente se gli autori utilizzassero annotazioni per tenere conto del grado di interesse dell'utente per ogni evento.

Lo sfruttamento dei comportamenti degli utenti soffre ancora di alcune limitazioni. (Hoebel et al, 2006) presentando il progetto Gugubarra 2.0 che è un miglioramento di Gubugarra 1.0 (Mushtaq et al, 2004). (Hoebel et al, 2006) ha illustrato alcuni problemi limitati, come la durata della visita di una pagina web, che è in qualche modo complicata perché non possiamo essere sicuri di ciò che l'utente sta esattamente facendo davanti al suo schermo, così in (Hoebel et al, 2006) i ricercatori hanno permesso ai proprietari dei siti web, nell'ambito dello scoop della profilazione dell'utente, di suddividere le pagine in zone e di chiedere agli utenti i loro interessi, ma la maggior parte delle volte gli utenti trovano fastidioso dover sempre rispondere a queste domande, così danno risposte che non riflettono i loro reali interessi.

I social media, i blog, i post, le conversazioni elettroniche (Nicoletti et al, 2012) e i messaggi di posta elettronica sono un'importante fonte di informazioni che trasmettono informazioni rilevanti sugli utenti. Per sfruttare questo tipo di dati, sono stati proposti diversi metodi per estrarre contenuti semantici dai testi. (Jusoh et al, 2009) hanno proposto un framework per l'estrazione di Entità Nominate (NE) utilizzando teorie della probabilità, regole grammaticali fuzzy e predicati per determinare la reale semantica di una frase. Vi sono una serie di metodi di estrazione delle informazioni utilizzati in diversi campi: riassunto del testo e Information Retrieval (IR), sfruttamento dei vantaggi combinando la potenza formale delle metriche statistiche, la potenza semantica delle ontologie e le tecniche di NLP per rilevare meglio le preferenze degli utenti, i dati personali e professionali, ecc, proponendo PBS (Profile Builder System) per la costruzione dei profili degli utenti a partire dalle domande di lavoro via e-mail, nell'ambito dell'e-recruitment.

In alcuni casi, le fonti di informazione sono limitate, come ad esempio Twitter, dove i tweet degli utenti sono limitati a 140 caratteri (Michelson et al, 2010), i titoli dei blog (Oka et al, 2006), ecc, quindi la loro semantica è difficile da individuare.

(Abel et al, 2011), ispirandosi all'osservazione di (Kwak et al, 2010) secondo cui l'85% dei tweet è legato alle notizie, hanno proposto un approccio per arricchire i profili degli utenti utilizzando la libreria di (Kohlschutter et al, 2010) e i servizi web forniti da OpenCalais⁷ per estrarre meglio i NE, identificando organizzazioni, eventi e argomenti di interesse veicolati dai tweet degli utenti.

1.4.3 Gestione del profilo utente

La diversità di sistemi, servizi e applicazioni permette agli utenti di interagire in modo diverso, consentendo loro di esprimersi attraverso profili diversi, con informazioni a volte ridondanti e a volte incomplete. Per arricchire i profili degli utenti e facilitare le loro interazioni con i vari sistemi informativi è necessario un meccanismo di gestione.

⁷ <http://www.opencalais.com>

La gestione dei profili degli utenti si trova ad affrontare molte sfide tecniche utilizzando diversi metodi, tra cui tecniche di costruzione e modellazione (Figura 1). Uno degli obiettivi più importanti di questo asse è garantire l'interoperabilità tra le diverse applicazioni e i servizi di ambienti diversi. L'interoperabilità è la capacità di due o più sistemi o componenti di scambiare informazioni e di utilizzare le informazioni scambiate (Geraci, 1991). Si riferisce alla capacità di accedere e interpretare informazioni derivate da più fonti eterogenee. Può essere

- (i) sintattica, dove le differenze tra i modelli sono risolte a livello di applicazione, o
- (ii) semantica, dove le differenze tra i modelli sono risolte a livello di conoscenza (Carmagnola e Cena, 2009).

(Wagner et al, 2013) ha proposto un gestore di profili ubiqui *eProfile*, di dominio generico orientato ad ambienti distribuiti, con interoperabilità tra sistemi.

eProfile, (Figura 1) deduce i dati del profilo attraverso l'analisi delle tracce di un'entità (fornite da un gestore di tracce, che restituisce tutti gli eventi posseduti da un'entità), seguendo le regole definite dalle entità, e fornisce i dati inferiti per essere utilizzati dalle applicazioni.

Le caratteristiche principali di *eProfile* sono l'uso delle tracce per estrarre i profili, la capacità di generare profili dinamici, la capacità di gestire regole e modelli di inferenza dinamici e l'operatività semantica del modello, dal momento che *eProfile* supporta più applicazioni. Come esempio di applicazione, permette di mantenere aggiornati i profili dinamici degli studenti, consentendo un migliore adattamento all'ambiente di apprendimento ubiquo.

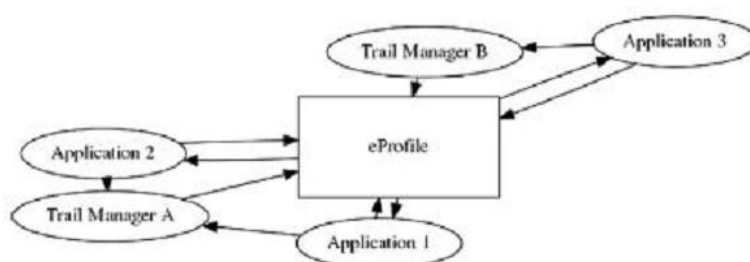


Figura 1. *eProfile* collegato a più gestori di percorsi. (Wagner et al, 2013)

Alcune altre problematiche sono illustrate in (Ghosh et al, 2009) (Ghosh et al, 2008) come:

- Scarsità del profilo, nota come problema dell'avvio a freddo: poiché gli utenti non amano ripetere le loro informazioni per ogni servizio con cui interagiscono, in alcuni casi, quando interagiscono per la prima volta con un nuovo servizio, questo ha informazioni estremamente scarse su di loro.
- Personalità dell'utente, legata all'influenza del contesto e dell'ambiente sui suoi comportamenti e sulle sue preferenze.
- Utilizzo di dati fuori banda, in altre parole, il problema dell'utilizzo di frammenti diversi di interazioni o preferenze dell'utente provenienti da diversi servizi e applicazioni.
- Sovraccarico dei server dovuto alla memorizzazione di diversi profili appartenenti allo stesso utente.

Molti lavori e strumenti sono stati realizzati e sviluppati per rispondere alle sfide di cui sopra, come ad esempio: Google OpenSocial⁸ e Yahoo Y!OS⁹.

⁸ Google OpenSocial, code.google.com/opensocial

⁹ Yahoo Y!OS, developer.yahoo.com/yos

Dai lavori analizzati, si possono notare diverse tecniche di gestione, come il matching, l'aggregazione, la negoziazione e la selezione, ecc., in cui l'obiettivo è selezionare il profilo o l'insieme di profili più adeguato a una determinata situazione.

Il concetto di **aggregazione** si ritrova in molti lavori, come ad esempio (Ghosh et al, 2009), in cui gli autori hanno proposto un framework per la gestione dei profili utente, che fa parte di HP Labs, InfoPal Profile Manager (IPM), che si occupa di gestire le informazioni sul profilo della banda esterna.

Con l'emergere di servizi che esportano i loro dati attraverso API standard (Ghosh et al, 2008), è possibile fornire strutture per scoprire e costruire dinamicamente un profilo utente adatto a partire da frammenti di profilo distribuiti tra più servizi. In questo lavoro gli autori utilizzano un agente per costruire un corpus di profili utente, mantenuti da diversi servizi, ciascuno descritto con una specifica ontologia.

Il mediatore calcola la distanza semantica tra i profili del corpus e le specifiche fornite dal servizio richiedente e, in base a una soglia, i profili vengono utilizzati per costruire un profilo utente per questo servizio. La possibilità di determinare quali informazioni sono più rilevanti risolve il problema della scarsità. Inoltre, IPM supporta lo *user profile personae*, consentendo agli utenti di memorizzare diversi *profile personae* e di scegliere quale *personae* inoltrare a un servizio. Lo svantaggio di questo approccio è che sono necessari schemi e accordi espliciti di condivisione dei dati punto-punto da parte dei diversi servizi.

All'interno dello stesso scoop, si trova (Morikawa et al, 2005) che propone una *Profile Aggregation* (PA) come nodo di raccolta di vari profili da diverse fonti e di gestione sotto forma di "profilo personalizzato" con un meccanismo di controllo della privacy in grado di verificare il tipo di dati che possono essere condivisi, con chi e in quali situazioni, utilizzando il *Resource Description Framework* (RDF) (Beckett, 2004) per una migliore semantica.

I social media sono un'altra area interessante, che ha un grande impatto sulle persone di tutto il mondo. Non è raro che gli individui creino più profili in diversi social network (Facebook, Twitter, LinkedIn, ecc.), ognuno dei quali contiene serie di informazioni personali parzialmente sovrapposte. La corrispondenza di questi diversi profili permette di creare un profilo globale che fornisce una visione olistica delle informazioni di un individuo. Come obiettivo la gestione delle informazioni dell'utente, in (Bennacer et al, 2014) i ricercatori hanno proposto un algoritmo per la ricerca di profili su n reti sociali, sfruttando i collegamenti tra i profili e i valori degli attributi del profilo, per determinare le coppie che corrispondono.

Ispirandosi alle osservazioni di Krishnamurthy et al, che hanno identificato un insieme di attributi comuni disponibili in 12 dei più importanti social network (Krishnamurthy et al, 2009) (come nomi, cognomi, username), gli autori in (Bennacer et al, 2014) si concentrano su username, nome, email e link ad altre pagine Web.

In ambito medico, per una migliore qualità dei servizi sanitari, (Fengou et al, 2012) utilizzano una cartella clinica elettronica (eMHR) per fornire un sistema ubiquo personalizzato. Diverse entità sono collegate al paziente (rete sociale del paziente, centro sanitario, casa intelligente, ufficio del paziente, pazienti, medici, famiglia del membro), al fine di fornire servizi sanitari personalizzati di alta qualità e affidabili al paziente quando le sue condizioni di salute sono critiche.

Ad esempio, ogni paziente è collegato a un gruppo di profili indicizzati relativi alle sue condizioni di salute; in un caso critico, viene creato un profilo sanitario dell'utente e vengono selezionati e consigliati il profilo del centro sanitario e del medico più adeguati.

In (Cherrat et al, 2015) i ricercatori hanno proposto un sistema di estrazione dei dati per la gestione del profilo dell'utente basato sul comportamento per sviluppare un sistema di sicurezza e sorveglianza per i

conducenti. Avendo una forte connessione tra stanchezza, sonno e comportamenti di guida scorretti ed errori che portano ad incidenti, gli autori hanno proposto un sistema intelligente per gestire tutti i comportamenti di guida scorretti per prevenire incidenti e situazioni di pericolo. Gli autori hanno utilizzato il metodo PCA (Freund e Schapire, 1996) e l'algoritmo di Viola e Jones (Viola e Jones, 2001) per rilevare gli oggetti nelle immagini digitali, sensori speciali per monitorare il conducente e il veicolo e servizi web per mettere in contatto i conducenti con gli agenti di polizia.

Nell'ambiente Cloud, il *Knowledge-as-a-Service*, secondo (Li et al, 2009) è indispensabile per l'apprendimento, il lavoro e la vita dell'uomo e il Cloud computing, come nuova tecnologia emergente, sta fornendo tutti i tipi di servizi scalabili che consentono la creazione, l'utilizzo e la condivisione della conoscenza personale, come proposto da (Tsui et al, 2011), un Cloud-Based Personal Knowledge Management (CBPLM).

(Jeong et al, 2012) nell'ambito dello scoop del mobile cloud computing propongono una piattaforma di intelligenza basata sul contesto per gestire le risorse utilizzando le informazioni del contesto. Propongono un processo di aggregazione delle informazioni utilizzando un'ontologia generica. Utilizzano inoltre vincoli di tempo e di posizione per la raccomandazione delle risorse.

1.5 Analisi critica

Nella Parte precedente si è cercato di illustrare i lavori più importanti nell'ambito della profilazione degli utenti, presentando ogni fase (modellazione, costruzione e gestione dei profili utente).

Sono stati annotati i vantaggi e i limiti da tenere in considerazione nel proporre un sistema di profilazione degli utenti. Ne sono emersi alcuni punti di discussione, quali:

- L'uso delle ontologie è indispensabile per il loro potere semantico, ma si deve anche dire che un profilo utente è più di un insieme di parole chiave o concetti (Iggui et al, 2016), per questo motivo è raccomandabile combinare le ontologie con misure statistiche e altri metodi formali.
- Alcune delle strutture di profilo proposte hanno lo svantaggio di essere limitate dai tipi di informazioni (attributi) definiti dal modello.
- Bisogna tenere in considerazione il contesto dell'utente e della sua dinamicità durante il processo di profilazione, in quanto importante fonte di informazioni in grado di orientare i risultati dei sistemi informativi. Va però precisato che il contesto dell'utente non è limitato alle sole informazioni sui dispositivi, come viene considerato da alcuni lavori.
- La posizione dell'utente, una componente del contesto dell'utente, è raramente presa in considerazione, mentre può guidare le preferenze, il tipo di ricerche e persino le attività di un dato utente, ad esempio non si fanno le stesse cose in ufficio e in un caffè.
- Le informazioni sui profili degli utenti devono essere sempre accessibili per l'aggiornamento, al fine di mantenere la freschezza dei dati e garantire una migliore qualità dei servizi.
- Gli utenti tendono a preferire approcci non supervisionati, per questo motivo l'"interoperabilità" è uno dei vincoli più importanti da considerare durante la modellazione o la gestione dei profili utente, al fine di esportare e sfruttare il profilo costruito da parte di diversi servizi e applicazioni. È stato citato XML come una delle tecnologie standard utilizzate per l'interoperabilità da diversi lavori.
- Come ultimo punto, non si può negare il fatto che ogni utente può avere diversi profili con informazioni diverse, a volte incomplete, in alcuni casi complementari in altri ridondanti. La ridondanza è una sfida molto importante per molti ricercatori, specialmente in ambiente Cloud, dove ogni dato è replicato su diversi server (problema del sovraccarico dei server).

Per arricchire il profilo dell'utente, per facilitare le sue interazioni con i diversi sistemi informativi (per non fornire ogni volta le sue informazioni), per minimizzare il sovraccarico dei server, per proporre una soluzione al problema del cold start e come frutto della nostra attuale indagine, si propone di seguito un Sistema di Gestione dei Profili Utente che ha come obiettivo la costruzione di un Profilo Utente Globale (UPMS-CGUP) a partire da un insieme di profili utente locali appartenenti a social Cloud.

1.6 Gestione dei profili utente, cloud e big data

1.6.1 Cloud e Big Data

Il cloud computing è un nuovo paradigma informatico, di cui non esiste ancora una definizione comune (Voas e Zhang, 2009).

Qui si adotta la definizione del National Institute of Standards and Technology (NIST) (Mell et Grance, 2011): Il cloud computing è un modello che consente l'accesso alla rete in modo ubiquo, conveniente e su richiesta a un pool condiviso di risorse informatiche configurabili (ad esempio, reti, server, storage, applicazioni e servizi) che possono essere rapidamente messe a disposizione e rilasciate con un minimo sforzo di gestione o di interazione con il fornitore di servizi.

Con l'avvento di Internet e l'esplosione delle fonti di dati, sempre più aziende si trovano ad affrontare nuove sfide che prima erano molto rare. (Sakr et al, 2011) presenta un'indagine completa sui meccanismi di gestione dei dati, sui problemi aperti e sulle sfide future in ambiente Cloud. Uno di questi problemi riguarda l'archiviazione e la fruibilità di insiemi di dati così grandi e spesso in rapida crescita, che gli strumenti abituali non sono più adatti. (Thibault, 2011) ha discusso il problema dell'archiviazione dei dati e dell'elasticità dei database Cloud presentando un confronto tra i *Relational Database Management System* e i database NoSQL, e come questi ultimi affrontino le nuove sfide. Questi nuovi, enormi insiemi di dati sono comunemente chiamati "Big Data"; questi dati possono appartenere a diversi servizi e applicazioni, emersi soprattutto con il Cloud computing. In questa sezione, i dati si riferiscono alle informazioni degli utenti. Gli utenti interagiscono con diversi servizi Cloud che richiedono informazioni diverse, generando così diversi profili locali su diverse applicazioni, appartenenti allo stesso utente. Facendo l'esempio dei social media (Facebook, Twitter, LinkedIn, ecc.), un utente può avere diversi profili su ogni applicazione, con informazioni diverse e/o ridondanti. Per ovviare, principalmente, ai problemi di cold start, interoperabilità, archiviazione dei dati e sovraccarico dei server, per una maggiore sicurezza dei dati, per una migliore gestione delle identità digitali, per facilitare i processi di autenticazione e per soddisfare meglio le esigenze degli utenti, di seguito si propone una piattaforma di gestione dei profili utente.

1.6.2 Servizio di gestione dei profili utente

La diversità dei servizi e delle applicazioni permette agli utenti di interagire quotidianamente con diversi sistemi, esprimendosi in modo diverso attraverso vari profili con informazioni ridondanti e talvolta incomplete.

La gestione dei profili degli utenti è un meccanismo molto importante, per identificare le informazioni rilevanti di un individuo e facilitare le sue interazioni con i diversi sistemi informativi (per non fornire ogni volta le sue informazioni), per migliorare i risultati dei diversi servizi e per una migliore qualità della personalizzazione, per arricchire il profilo dell'utente basandosi sulle diverse interazioni di questo con diversi sistemi e applicazioni, per ridurre il numero di profili appartenenti allo stesso utente, affrontando così il problema della ridondanza e del sovraccarico dei server.

Per affrontare il problema dell'avvio a freddo e per consentire l'interoperabilità tra le applicazioni, si propone un UPMS (User Profiles Management Service).

Traendo ispirazione da (Bennacer et al, 2014) e (Kostadinov, 2005), dove sono stati proposti metodi di matching semantico, si è scelto il social Cloud (social media) come esempio, per proporre una piattaforma di gestione dei profili utente che abbia come obiettivo la costruzione di un profilo utente globale.

Il profilo utente globale includerà tutte le informazioni appartenenti a un determinato utente. Sarà memorizzato a livello di Cloud e quindi accessibile da tutte le applicazioni e i servizi Cloud.

L'architettura corrispondente del sistema proposto è mostrata nella Figura 2, utilizzando l'architettura HP Cloud:

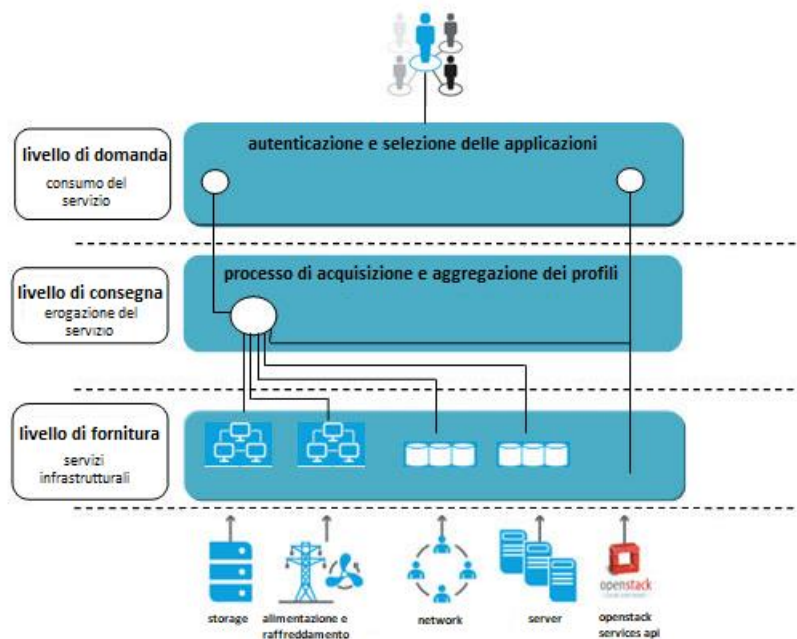


Figura 2. Servizio di gestione dei profili utente (UPMS) presentato sull'architettura HP Cloud.

Il presente approccio si basa sullo sfruttamento dei profili degli utenti locali nelle diverse applicazioni dei social media.

Di seguito, si illustrano in dettaglio le diverse fasi dell'approccio proposto:

Ipotesi: Si supponga che: Tutti i profili degli utenti siano stati creati con lo stesso indirizzo e-mail e la lingua utilizzata sia "inglese".

- 1) **Autenticazione e selezione dell'applicazione:** In base alla scelta dell'utente in questa fase, il sistema può essere supervisionato o meno:
 - Sistema supervisionato: per rispettare la privacy dell'utente, il sistema consente all'utente di scegliere le applicazioni da cui sarà costruito il suo profilo utente globale e di attribuire un grado di rilevanza a ciascuna di esse.
 - Sistema non supervisionato: in questo caso, il sistema prenderà in considerazione tutti i sotto-profilo relativi al dato indirizzo e-mail.
- 2) **Processo di acquisizione e aggregazione dei profili:** In questa fase il sistema ottiene l'accesso ai sotto-profilo appartenenti alle applicazioni scelte tramite le API corrispondenti. Una volta recuperati, vengono avviati i processi di allineamento semantico e di filtraggio.

- Allineamento semantico: consiste nello stabilire una corrispondenza semantica tra i lessici di due o più sotto-profilo utilizzando l'ontologia.
- Filtraggio delle informazioni: consiste nel selezionare le informazioni rilevanti (valori) in base alla rilevanza delle loro fonti e alla natura dell'attributo (monovalente o multivalente).

Bibliografia di riferimento

- (Jusoh et al, 2009): Jusoh, S., Fawareh, H. M. Semantic Extraction from texts. International Conference on Computer Engineering and Applications IPCSIT vol.2, Singapore, 2009.
- (McCallum, 2005): McCallum, A., Information extraction: distilling structured data from unstructured text, *ACM Queue*, vol. 3, no. 9, pp. 48–57, 2005.
- (Alfawareh et al, 2008): Alfawareh, H. M., Jusoh, S., and Osman, W. R. S., Ambiguity in text mining. *International Conference on Computer and Communication Engineering (ICCCCE'08)*, pp. 1172–1176, Kuala Lumpur, May 13-15, 2008.
- (Redfearn, 2006): Redfearn, J., Text mining, *JISC*, pp. 1–2. 2006
- (Schiaffino et al, 2009): Schiaffino, S., and Amandi, A., Intelligent User Profiling. *Artificial Intelligence*, LNAI 5640, pp. 193 – 216. Berlin Heidelberg, 2009.
- (Soniya et al, 2013): Soniya, M., and Kalpana, B., Performance Analysis of Ranking Algorithms Using Ontology for Web Personalization. *IJAIR*, Vol.2, Issue 3, 2013.
- (Kaczmarek et al, 2010): Kaczmarek, T., Zyskowski, D., Adam, W., and Witold, A, Information extraction from web pages for the needs of expert finding. *Studies in Logic, Grammar and Rhetoric*. Vol. 22(35), pp. 141-157, 2010.
- (Jeong et Choi, 2012): Jeong, H., Choi, E, User Authentication using Profiling in Mobile Cloud Computing. In proceeding (2) AASRI Conference on Power and Energy Systems. Pp. 262 - 267. Hong Kong, Settembre, 2012.
- (Salter et al, 2010): Satler, M., F., Romero, F., P., Menendez, V., H., Zapata, A., and Prieto, M., E, A Fuzzy Ontology Approach to represent User Profiles in E-Learning Environments. *International Conference Fuzzy Systems (FUZZ)*, IEEE, Barcelon, July, 2010.
- (Alguliev et al, 2014): Alguliev, R., Abdullaeva, F., User Profiles and Identifying User Behaviour in the Cloud Computing Environment. *Universal Journal of Communications and Network*, vol. 2, no. 4, pp. 63-68. 2014.
- (Kanoje et al, 2014): Kanoje, S., Girase, S., Mukhopadhyay, D., User Profiling Trends, Techniques and Applications. *International Journal of Advance Foundation and Research in Computer (IJAFRC)*, vol. 1, no. 1, 2014.
- (Bedekar et al, 2008) : Bedekar, M.V., Deshpande, B., Joshi, R., Web Search personalization by User Profiling. In *Emerging Trends in Engineering and Technology ICETET'08*, Nagpur, India, 2008.
- (Ouanaim et al, 2010): Ouanaim, M., Harroud, H., Berrado, A., Boulmalf. M., Dynamic User Profiling Approach for Services Discovery in Mobile Environments. *Proceedings of the 6th International Wireless Communication and Mobile Computing Conference IWCMC'10*, pp. 550-554, New York, 2010.
- (Bradley et al, 2000): Bradley, K., Rafter, R., Smyth, B., Case-Based User Profiling for Content Personalization. *International Conference Adaptive Hypermedia and Adaptive Web-Based System*, pp. 62-72, Berlin, 2000.
- (Aquino and Filgueiras, 2005): Aquino, P.T.J., Filgueiras, L.V.L., User Modeling with Personas. *Latin American Conference on Human-Computer Interaction CLIHC'05*, pp. 277-282, Cuernavaca, Mexico, October 23-26, 2005.
- (Iggui et al, 2016): Iggui, T., Nacer, H., Sklab, Y., Ait Radi, T., *Web Application for User Profiling*. *International Journal of Information Systems in the Service Sector, IJISSS*, vol. 8, no. 2, pp. 44-56, April – June, 2016.
- (Salem and Rauterberg, 2004): Salem, B., Rauterberg. M., Multiple User Profiling Merging (MUPE): Key Challenges for Environment Awareness. *Lecture Notes in Computer Science*, vol. 4161, pp. 103-116. 2004.
- (Iggui et al, 2018) : Iggui T., Nacer H., Tari A., Modeling, Building and Management User Profiles: UPMS (User Profiles Management Service). Published in *Proceedings of the International Conference on Industrial Engineering and Operations Management Paris, France, July 26-27, 2018*

- (Liu et al, 2009): Liu, H., Salem, B., Rauterberg, M., A Surevey on User Profile Modeling for Personalized Service Delivery Systems. Proceeding of IADIS International Conference on Interfaces and Human Computer Interaction, pp. 45-51, 2009.
- (Kay, 2001): Kay, K., User Modeling for Adaptation. User Interfaces for All: Concepts, Methods and Tools (Ed. Constantine Stephanidis), pp. 271-294, 2001.
- (Kuflik et al, 2012): Kaflik, T., Kay, J., Kummerfeld, B., *Challenges and Solutions of Ubiquitous User Modeling*. In Kruger, A. and Kuflik, T. (eds.), Ubiquitous Display Environments, Cognitive Technologies, Berlin, 2012.
- (Manning et al, 2008): Manning, C.D., Raghavan, P., Schutze, H., Introduction to Information Retrieval. Cambridge University Press, 2008.
- (Hafez and Xu, 2013): Hafez, A. A., Xu, Y., *A Survey of User Modelling in Social Media Websites*. Computer and Information Science, vol. 6, no. 4. September, 2013.
- (Alguliev et al, 2014): Alguliev, R., Abdullaeva, F., User Profiles and Identifying User Behaviour in the Cloud Computing Environment. Universal Journal of Communications and Network, vol. 2, no. 4, pp. 63-68. 2014.
- (Tayari et al, 2009): Tayari, I., Thanh, N. L., Ben Amar, C., Modélisation des Etats Emotionnels par un Espace Vectoriel Multi-dimensionnel. Laboratoire d'Informatique de Signaux et Systèmes de Sophia Antipolis – UNSA – CNRS, Décembre 2009.
- (Abel et al, 2011): Abel, F., Gao, Q., Houben, G.J., Tao, K., Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web. 8th Extended Semantic Web Conference, ESWC 2011, pp 375-389, Greece, May 29 – June 2, 2011.
- (EARL, 2006): Humaine Emotion Annotation and Representation Language (EARL), 30 June 2006.
- (Gruber, 2008): Gruber, T., What is an Ontology, Encyclopedia of Database Systems, vol. 1, 2008.
- (Daoud and Tamine, 2008): Daoud, M., Tamine-Lechani, L., Construction des profils utilisateurs à base d'ontologie pour une recherche d'information personnalisée. Journal of Laboratoire IRT University of Paul Sabatier, 2008.
- (Wasim et al, 2011): Wasim, M., Shahzadi, I., Ahmad, Q., Mahmood, W., Extracting and Modeling User Interests based on Social Media. Multitopic Conference (INMIC), IEEE 14th International, pp. 284-289. Karachi, 22-24 December. 2011.
- (Michelson et al, 2010): Michelson, M., Macskassy, S.A.. Discovering Users' Topics of Interest on Twitter: A First Look. The 4th workshop ACM on Analytics for Noisy unstructured text Data, pp.73-80, New York, USA. 2010.
- (Doan et al, 2002): Doan, A., Madhavan, J., Domingos, P., Halevy, A., Learning to Map between Ontologies on the Semantic Web, 11th International Conférence World Wide Web (WWW '02), pp. 662-673, 2002.
- (Han et Chang, 2002): Han. J., Chang, K.C.C., Data Mining for Web Intelligence, Computer, vol. 35, no. 11, pp. 64-70.2002
- (King et al, 2007): King, J.D., Li, Y., Tao, X., Nayak, R., Mining World Knowledge for Analysis of Search Engine Content, Web Intelligence and Agent Systems, vol.5, no.3, pp.233-253, 2007
- (Umamaheswari et Patil, 2012): Umamaheswari, B., Patil, P., Personalized Ontology Model – Survey. The 12th International Conference on Hybrid Intelligent System (HIS), pp. 348-353, India, 4-5 December 2012.
- (Strang et al, 2004): Strang, T., Linnhoff-Popien, C., A context modeling survey. Workshop on Advanced Context Modeling, Reasoning and Management associated with the 6th International Conference on Ubiquitous Computing (UbiComp), Nottingham, 2004.
- (Couto et al, 2005): Couto, R., Rocha, A., Endler, M., Evolutionary and efficient context management in heterogeneous environments. Proceeding of the 3rd International workshop on Middleware for Pervasive and Adhoc computing, pp. 1-7, New York, USA, 2005.
- (Schmohl & Baumgarten, 2008): Schmohl R., Baumgarten, U., Context-aware computing : a survey preparing a generalized approach. International Multi-Conference of Engineers and Computer Scientists, pp. 744-750, Hong Kong, March 2008.
- (Bouzeghoub et Kostadinov, 2005): Bouzeghoub M., Kostadinov D., Personnalisation de l'information: aperçu de l'état de l'art et définition d'un modèle flexible de profils, Conférence en Recherche d'Informations et Applications Complexe de Recherche Interprofessionnel en Aérothermochimie (CORIA), pp. 201-218, France 2005.
- (Kostadinov, 2005) : Kostadinov. D., Personnalisation de l'information : une approche de gestion de profils et de reformulation de requêtes. Human-Computer Interaction. Université de Versailles, France, 2007.

- (Sklab et al, 2013) : Sklab, Y., Nacer, H., Tari, A. K, Modeling User Context for Pervasive Computing. Proceedings of the International Workshop on Requirements Engineering, pp. 40-46, Constantine, Algeria, 2013.
- (Baldauf et al, 2007): Baldauf, M., Dustdar, S., A survey on context-aware system. International Journal of Ad Hoc and Ubiquitous Computing, vol. 2, pp.263_277, 2007.
- (Chaari, 2007) : Chaari, T., Adaptation d'applications pervasives dans des environnements multi-contextes, Laboratoire d'Informatique en Image et Systèmes d'information (LIRIS), September 2007.
- (Dey et al, 1999): Dey, A. K., Salber, D., Futakawa, M., et al. An architecture to support context aware applications. Technical report, ACM Symposium, <http://smartech.gatech.edu/jspui/bitstream/1853/3390/1/99-23.pdf> , June 1999.
- (Kankainen and Parkinen, 2001): Kankainen, T., Parkinen, J., GUP-Graphical Presentation of User Profile. In CHI Conference on Human Factors in Computing Systems, pp. 105-106, Seattle, Washington USA, 2001.
- (Jonassen et al, 1993): Jonassen, H. D., Beissner, K., Yacci, M. Structural Knowledge, Techniques for representing, conveying and Acquiring Structural Knowledge, Lawrence Erlbaum Associates, New Jersey, 1993.
- (Malone et al, 1987): Malone, T., Grant, K., Turbak, F., Brobst, S., Cohen, M., Intelligent information sharing systems. Communications of the ACM , vol. 30, no. 5, pp. 390-402, 1987.
- (Strang et al, 2004): Strang, T., Popien, C. L., A context modeling survey. In Proceedings of the 1st International Workshop on Advanced Context Modeling, Reasoning and Management, pp. 33-40, UK, 2004
- (Vanathi et al, 2011): Vanathi, B. V., Uthariaraj, R., Context aware system for smart consignment tracking. European Journal of Scientific Research, vol. 50, pp.586-596, 2011.
- (Poo et al, 2003): Poo, D., Ching, B., Goh, J.M., A hybrid approach for user profiling. Proceedings of the 36th Annual Hawaii International Conference on System Sciences, vol. 7, pp. 9-13, 2003.
- (Jie et al, 2010) : Jie, T.L.Y., Zhang, D., Zhang, J., A Combination Approach to Web User Profiling, ACM Transactions on Knowledge Discovery from Data, vol. 5, no. 1, Article 2, New York USA, December 2010.
- (Li et al, 2014) : Li, J., Alan, R., Eduard, H., Weakly Supervised User Profile Extraction from Twitter, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics ACL, pp. 165–174, Baltimore, USA, June 23-25 2014.
- (Hijikata, 2004): Hijikata, Y., Implicit user profiling for on demand relevance feedback. Proceedings of the 9th international conference on Intelligent user interfaces, pp. 198-205, Portugal 2004.
- (Stuart et al, 2003): Stuart, M.E., Shadbolt, N. R., Roure, D. C. D., Capturing interest through inference and visualization: Ontological user profiling in recommender systems. In Proceedings of the 2nd international conference on Knowledge capture, pp. 62-69. USA, 2003.
- (Shmueli-Scheuer et al, 2010) : Shmueli-Scheuer, M., Roitman, H., Carmel, D., Mass, Y., Konopnicki, D., Extracting User Profiles from Large Scale Data. Proceedings of the 2010 Workshop on Massive Data Analytics on the Cloud. Raleigh, North Carolina, USA. April 26, 2010. (Kullback et Leibler, 1951): Kullback, S., Leibler, R. A., On information and sufficiency. The Annals of Mathematical Statistics, vol. 22(1), pp. 79–86, 1951.
- (Widiantoro et Yen, 2001): Widiantoro, D.H., Yen, J., Incorporating Fuzzy Ontology of Term Relations in a Search Engine. In BISC International Workshop on Fuzzy Logic and the Internet. Berkeley: University of California, 2001.
- (Lau et al, 2009): Lau, R. Y. K., Song, D., Li, Y., Cheung, T. C. H., Hao, J. X., Toward a Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning. IEEE Transactions on Knowledge and Data Engineering, vol. 21(6), pp. 800-813, 2009.
- (Han and Chen, 2009): Han, L., Chen. G., A Fuzzy Clustering Method of Construction of Ontology-based User Profiles. Advances in Engineering Software, vol. 40(7), pp. 535-540, 2007.
- (Sieg et al, 2007) : Sieg, A., Mobasher, B., Burke, R., Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search. IEEE Intelligent Informatics Bulletin, vol.8, no.1. November 2007.
- (Morita and Shinoda, 1994) : Morita, M., Shinoda, Y., Information filtering based on user behavior analysis and best match text retrieval. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 272-281. Ireland, 1994.
- (Oard and Kim, 1998): Oard, D.W., Kim, J., Implicit feedback for recommender systems. Proceedings of the AAAI Workshop on Recommender Systems. 1998.

- (Xing et al, 2011): Xing, K., Zhang, B., Zhou, B., Liu, Y., School of Computer Engineering & Science Shanghai University, Shanghai, China. Behavior based User Interests Extraction Algorithm. IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing, 2011.
- (Stupar et al, 2014): Stupar, G., Nad, D., Beserminji, A., Papp, I., One solution of STB Users Cloud Based Profiling. Published in Proceeding of the 4th IEEE International Conference on Consumer Electronics (ICCE), pp. 364-366, Berlin. 7-10 September 2014.
- (Hoebel et al, 2006): Hoebel, N., Kaufmann, S., Tolle, K., Zicari, R.V., The Design of Gugubarra 2.0: A Tool for Building and Managing Profiles of Web Users. Published in the proceedings of the International Conference on Web Intelligence IEEE/WIC/ACM (WI'06), pp. 317-320. Hong Kong, 18-22 December 2006.
- (Hoebel et al, 2006): Hoebel, N., Kaufmann, S., Tolle, K., Zicari, R., The Gugubarra Project: Building and Evaluating User Profiles for Vistors of Web Sites, First IEEE Workshop on Hot Topics in Web Systems and Technologies, Boston, Massachusetts, USA, November 2006.
- (Mushtaq et al, 2004): Mushtaq, N., Tolle, K., Werner, P., Zicari, R., Building and Evaluating Non-Obvious User Profiles for Visitors of Web Sites, IEEE Conference on E-Commerce Technology (CEC 04), San Diego, California, USA, 2004.
- (Nicoletti et al, 2012): Nicoletti, M., Schiaffino, S., Godoy, D., Mining interests for user profiling in electronic conversations. Journal of Isistan Research Institute, Conicet-Unicien, University of Paraje Arroyo Seco, 2012.
- (Oka et al, 2006): Oka, M., Abe, H., Kato, K., Extracting Topics From Weblogs Through Frequency Segments. Workshop of WWW'2006. Edinburgh, UK. May 22-26, 2006.
- (Kwak et al, 2010): Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? The 19th ACM International Conference on WWW, pp. 591-600, New York, NY, USA, 2010.
- (Kohlschütter et al, 2010): Kohlschütter, C., Fankhauser, P., Nejd, W., Boilerplate detection using shallow text features. 3rd ACM International Conference on Web Search and Data Mining. pp. 441-450, New York, 2010.
- (Geraci, 1991): Geraci, A., IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries, New Jersey, USA, 1991.
- (Carmagnola & Cena, 2009): Carmagnola, F., Cena, F., User identification for cross-system personalization, Journal of Information Sciences, vol. 179, no. 1, pp. 16-32, 2009.
- (Wagner et al, 2013): Wagner, A., Barbosa, J. L.V., Barbosa, D.N.F., A model for Profile Management Applied to Ubiquitous Learning Environments. In Expert Systems with Applications, vol. 41, pp. 2023-2034, 2014.
- (Ghosh et al, 2009): Ghosh, R., Dekhil, M., Discovering Profiles. The 18th International Conference on World Wide Web, pp. 1233-1234, Madrid, Spain, April 20-24, 2009.
- (Ghosh et al, 2008): Ghosh, R., Dekhil, M., Mashups for semantic user profiles. The 17th International Conference of World Wide Web, pp. 1229-1230, Beijing, 2008.
- (Morikawa et al, 2005): Morikawa, D., Honjo, M., Yamaguchi, A., Ohashi, M., A Proposal of User Profile Management Framework for Context-Aware Service. In Proceedings of the Symposium on Applications and the Internet Workshops (SAINT-W'05) IEEE Computer Society. Italy, February, 2005.
- (Beckett, 2004): Beckett, D., RDF/XML Syntax Specification, W3C, <https://www.w3.org/TR/rdf-syntax-grammar/> February, 2004.
- (Jeong et al, 2012): Jeong, H., Jeong, Y., Kang, Y., Kim, M., Choi, E., Efficient Method of Profiling on Mobile Cloud Computing. International Conference on Future Information Technology and Management Science & Engineering Lecture Notes in Information Technology, vol.14, pp. 177-182, 2012.
- (Tsui et al, 2011): Tsui, E., Cheong, R.K.F., Sabetzadeh, F., Cloud-Based Personal Knowledge Management as a service (PKMaaS). Published in International Conference on Computer Science and Service System (CSSS), pp. 2152-2155, Nanjing, 27-29 June, 2011.
- (Li et al, 2009): Li, L., Zheng, Y., Zheng, F., Zhong, S., Cloud Computing Support for Personal Knowledge Management, International Conference on Information Management, Innovation Management and Industrial Engineering, pp. 171-174. 2009.
- (Bennacer et al, 2014): Bennacer, N., Jipmo, C.N., Penta, A., Quercini, G., Matching User Profiles Across Social Network, Springer International Publishing Switzerland, pp. 424-438. 2014.
- (Krishnamurthy et al, 2009): Krishnamurthy, B., Wills, C.E.: On the Leakage of Personally Identifiable Information via Online Social Networks. The 2nd ACM Workshop on Online Social Networks, pp. 7-12, 2009.

- (Weingartner et al, 2015): Weingartner, R., Brascher, G.B., Westphall, C. B., Cloud resource management: A survey on forecasting and profiling models. *Journal of Network and Computer Applications*, no. 47, pp. 99-106, 2015.
- (Fengou et al, 2012): Fengou, M.A., Mantas, G., Lymberopoulos, D., Komninos, N., Ubiquitous Healthcare Profile Management Applying Smart Card Technology. *International Conference on Wireless Mobile Communication and Healthcare MobiHealth: Wireless Mobile Communication and Healthcare, LNICST*, vol. 83, pp. 248-255, 2012.
- (Cherrat et al, 2015): Cherrat, L., Ezziyyani, M., ElMouden, A., Hassar, M., Data Extraction for User Profile Management based on Behavior. *International Conference on Cloud Technologies and Applications, Morocco*, 2015.
- (Viola and Jones, 2001): Viola, P., Jones, M., Rapid object detection using a boosted cascade of simple features, In *CVPR*, pp. 511-518, USA, 2001.
- (Freund and Schapire, 1996): Freund, Y., Schapire, R.E., Experiments with a new boosting algorithm, In *Machine Learning*, pp.148–156, 1996.
- (Mell et Grance, 2011): Mell, P., Grance, T., The NIST definition of cloud computing, *National Institute of Standards and Technology*, pp. 800-145, Settembre 2011.
- (Voas and Zhang, 2009): Voas, J., Zhang, J., *Cloud Computing: New Wine or Just a New Bottle?*. Published by the IEEE Society. *Computer.org/ITPro*. 2009.
- (Sakr et al, 2011): Sakr, S., Liu, A., Batista, D. M., Alomari, M., A Survey of Large Scale Data Management Approaches in Cloud Environments. Published in *Communication Surveys & Tutorials, IEEE*, vol. 13, no. 3, pp. 311-336, 2011.
- (Thibault, 2011): Thibault. D., *Study and Comparison of Elastic Databases: Myth or Reality?*, Master degree in science computing and engineering ,2011.

2 Approcci alla profilazione degli utenti

Oggi, il rapido sviluppo delle informazioni ha causato un sovraccarico sul web, che aumenta la domanda di approcci personalizzati per trovare le informazioni richieste in modo dispendioso per gli utenti. La ricerca sulla personalizzazione è utilizzata in diverse applicazioni come la raccomandazione (ad es. di prodotti su piattaforme e-commerce), l'e-learning e il reperimento di informazioni personalizzate, ecc.

La profilazione di un utente Web è il processo di identificazione dei dati relativi al suo dominio di interesse che vanno a formare il cosiddetto "user model" [35].

Un profilo utente è generalmente utilizzato per la personalizzazione e la modellazione dell'utente. Riflette informazioni personali o fatti interessanti su un singolo utente, informazioni demografiche, ad esempio nome, età, livello di istruzione, paese, ecc. Dovrebbe anche catturare il comportamento dell'utente (argomenti interessanti, valutazioni, modelli, obiettivi, ecc.) quando egli interagisce con il Web.

La modellazione dell'utente è il processo di raccolta di informazioni sugli interessi di un utente, la costruzione, la manutenzione e l'utilizzo di profili utente. Ad esempio, nei sistemi di e-business, questo processo consente di catturare le caratteristiche degli utenti online, di trovare utenti online simili, di fornire prodotti e servizi personalizzati e quindi di migliorare la soddisfazione degli utenti.

Le tecniche di profilazione degli utenti hanno trovato ampia applicazione nella ricerca sul web, nei sistemi software adattivi all'utente, nell'identificazione degli utenti web, nella personalizzazione, nella raccomandazione, nell'analisi del mercato elettronico, nei sistemi di tutoraggio intelligente, negli agenti intelligenti, nonché nel recupero e nel filtraggio di informazioni personalizzate.

Il contenuto e la quantità di informazioni all'interno di un profilo utente possono variare a seconda dell'area di applicazione.

L'accuratezza del profilo utente si basa sul modo in cui le informazioni dell'utente vengono raccolte e organizzate. Inoltre, l'accuratezza di queste informazioni riflette l'utente durante la costruzione del profilo.

Il processo di modellazione dell'utente prevede due approcci: La generazione di un profilo utente iniziale per un nuovo utente e l'aggiornamento continuo delle informazioni del profilo per adattarsi al continuo cambiamento degli interessi, delle preferenze e delle esigenze dell'utente [14].

Una delle principali sfide nell'ambito della profilazione degli utenti è il modo in cui il profilo utente può essere costruito per riflettere accuratamente le preferenze degli utenti.

Questa Sezione esamina le sfide dell'attuale struttura di personalizzazione, che dipende per la maggior parte da profili statici o dinamici di basso livello e non è in grado di rappresentare la multidimensionalità dei profili degli utenti. Una personalizzazione statica porta a raccomandare servizi irrilevanti quando le esigenze e gli interessi dell'utente cambiano, perché utilizza le stesse informazioni dell'utente nel tempo [37].

Un'altra sfida consiste nell'affrontare il problema della partenza a freddo nei sistemi di raccomandazione e nel raccogliere implicitamente le informazioni con un'elevata accettazione del modello basato sulla fiducia. Per superare queste sfide, è necessario costruire un modello di utente che catturi e apprenda dinamicamente gli interessi e le preferenze dell'utente; per questo sono stati utilizzati vari algoritmi di clustering e classificazione per creare profili utente più completi.

In questa Sezione si esplorano alcuni aspetti del sistema di profilazione dell'utente disponibili nella letteratura esistente e si propone uno schema di classificazione innovativo per la ricerca sulla costruzione e la personalizzazione del profilo dell'utente. Lo schema di classificazione proposto comprende tre fasi principali: il processo di raccolta dei dati, la costruzione del profilo utente e le applicazioni personalizzate.

Questa Sezione è organizzata come segue: La Parte 1 presenta lo schema di classificazione proposto per il lavoro di ricerca sulla profilazione dell'utente, mentre le Parti 2, 3 e 4 illustrano gli sforzi di ricerca rispettivamente per la raccolta dei dati, la costruzione e la modellazione di un profilo utente e la personalizzazione.

Le principali lacune riscontrate negli attuali lavori di ricerca sono rappresentate nella Parte 5.

2.1 Schema di classificazione della profilazione degli utenti

In questa sezione si illustra uno studio completo delle ricerche recenti sulla profilazione degli utenti. La Figura 3 presenta lo schema di classificazione proposto per gli studi sulla profilazione degli utenti. Il processo di modellazione studia innanzitutto le informazioni che costituiscono il profilo dell'utente e la raccolta dei dati grezzi sull'utente e il modo in cui possono essere estratti. In secondo luogo, vengono studiati i processi di costruzione e mantenimento, le modalità di rappresentazione di un modello utente e le tecniche di costruzione. In terzo luogo, come identificare l'utente e come esplorare le applicazioni nel contesto della profilazione dell'utente al fine di fornire servizi personalizzati.

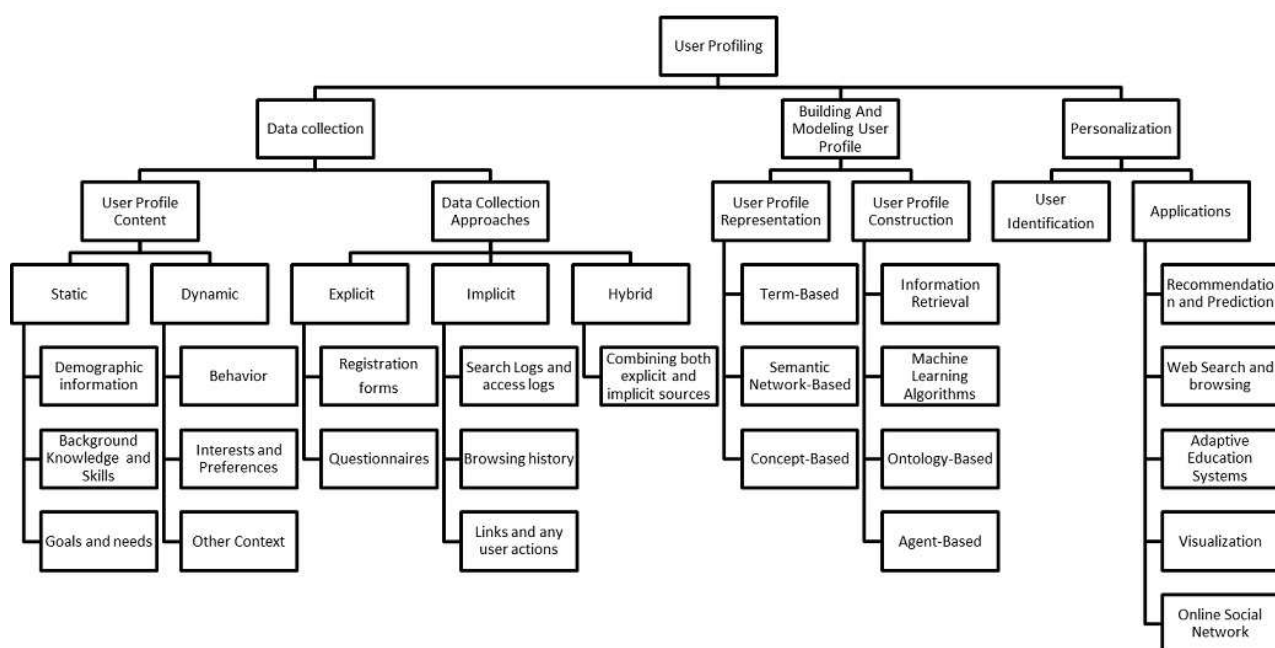


Figura 3. Schema di classificazione per la profilazione degli utenti

2.2 Raccolta dei dati

La fase principale della profilazione dell'utente consiste nel raccogliere dati sull'utente. Ottenere ed estrarre i dati dell'utente e conoscere i suoi interessi è un compito fondamentale per la creazione di sistemi personalizzati. Per questo motivo, la natura dei dati accessibili per un sistema personalizzato determina il modo e gli elementi di raccomandazione attraverso i quali il sistema può eseguire la personalizzazione.

2.2.1 Contenuto del profilo utente

Esistono diversi tipi di dati che possono essere presenti in un profilo utente: dati personali dell'utente, conoscenze e competenze di base dell'utente, esigenze e obiettivi dell'utente, che sono tipi di dati statici [4]. In alternativa, il comportamento dell'utente, i suoi interessi e preferenze e il contesto sono dinamici. Le seguenti sottosezioni sono suddivise in base alla dinamicità dei dati, da quelli di basso livello a quelli dinamici di alto livello.

2.2.1.1 Dati demografici. Le caratteristiche demografiche di un utente sono le caratteristiche di base, come il nome, il Paese, il sesso, l'età o la lingua madre, l'istruzione, i membri della famiglia, ecc.

Prajakta et al. [30] hanno confrontato alcuni framework che utilizzano diversi approcci di data mining e machine learning per la modellazione e la previsione dei dati demografici.

Fernandez et al. [7] hanno proposto un modello che cattura le informazioni demografiche sull'utente e le sue azioni all'interno delle comunità online.

Pazzani et al. [25] hanno raggruppato gli utenti in base ai loro dati demografici (come la posizione, il lavoro, la lingua, ecc.) per costruire un classificatore per gli studenti.

2.2.1.2 Background, conoscenze e abilità. Questi aspetti sono particolarmente importanti per la modellazione degli studenti nei sistemi di tutoraggio e nei sistemi educativi adattivi.

La conoscenza dello studente è la caratteristica principale che definisce un sistema adattivo e la sua conoscenza della materia insegnata è importante per fornire un'assistenza adeguata allo studente o per adattare il contenuto dei corsi in base ad essa [8].

Sistemi diversi gestiscono in modo diverso la sfida delle conoscenze inesatte o delle idee errate. In alcuni sistemi di raccomandazione, le competenze di un utente o di un dipendente, i ruoli che ricopre all'interno di un'organizzazione e le sue prestazioni in questi ruoli sono necessari per fornire le mansioni più adatte al candidato e raccomandare alcuni candidati idonei a un selezionatore [12].

2.2.1.3 Obiettivi e bisogni. Gli obiettivi di un utente sono importanti per individuare il suo scopo dall'applicazione che sta utilizzando, cioè ciò che l'utente vuole ottenere. Il bisogno di informazioni è l'obiettivo di ottenere informazioni rilevanti quando l'utente naviga sul Web.

È fondamentale imparare a prevedere gli obiettivi in alcuni sistemi di dialogo intelligenti. La modellazione è vicina alle dimensioni del concetto di bisogni, esigenze e desideri (needs, requirements and desires - NRD), che è stato applicato nei sistemi di fornitura di servizi personalizzati per adattare il comportamento degli utenti.

Periklis et al. [35], hanno discusso che gli interessi dell'utente, specificati esplicitamente sotto forma di parole chiave, possono essere trattati come obiettivi da raggiungere durante una sessione di navigazione; questi obiettivi sono memorizzati sul lato server.

2.2.1.4 Comportamento. Il comportamento è un tipo di informazione che viene raccolta implicitamente. È importante che il comportamento sia ripetitivo per rilevare i modelli utilizzati da un sistema adattivo o da un agente intelligente per adottare un sito web o per assistere il comportamento dell'utente.

Jie et al. [36] hanno utilizzato tre tipi di comportamenti dell'utente: tempo di attenzione, click e movimento del mouse come potenziali indicatori degli interessi degli utenti per rappresentare le preferenze degli utenti come feedback implicito.

Altri sistemi, come quello di [30], tengono traccia solo del comportamento degli utenti che cliccano, che comprende le query inviate e i documenti cliccati.

Una strategia adattativa simile è stata proposta anche sulla base dei comportamenti degli utenti [2],[13].

2.2.1.5 Interessi e preferenze. Gli interessi e le preferenze di un utente sono la parte principale della personalizzazione. Gli interessi possono significare argomenti sociali, argomenti di siti web, argomenti di documenti o argomenti di lavoro. Questi argomenti possono essere ricavati da varie fonti, come la cronologia degli acquisti o la cronologia di navigazione.

A volte gli interessi degli utenti sono classificati come interessi a breve o a lungo termine [15]. Alcuni framework eseguono la personalizzazione sulla base di entrambi, interessi a lungo termine come preferenze e interessi a breve termine.

Li et al. [17] hanno modellato rappresentazioni diverse per le attività frequenti e a lungo termine dell'utente in base alle sue preferenze. I profili a lungo termine sono costruiti a partire dalla directory di Google, utilizzando gli argomenti dei risultati dei flussi cliccati, mentre i modelli a breve termine utilizzano una cache di risultati cliccati di recente.

2.2.1.6 Altro contesto. Il contesto è identificato come qualsiasi informazione utilizzata per caratterizzare lo stato di qualsiasi elemento.

Jie et al. [36] hanno utilizzato un approccio di machine learning supervisionato per estrarre un profilo utente da social media come Twitter. Hanno discusso alcuni sistemi adattivi basati sul contesto dell'utente, in cui le informazioni vengono visualizzate in base alla posizione corrente dell'utente. Una modellazione del profilo utente basata sulle interazioni degli utenti con i social network, i tag e le annotazioni nei sistemi di social bookmarking e i post sul Web, come le connessioni seguite/followed in Twitter [18].

Anche le conoscenze su un utente, tra cui i suoi gusti, le sue antipatie e persino il suo stato emotivo, possono essere determinate utilizzando la profilazione. Le emozioni di un utente hanno un impatto diretto sul modo in cui utilizza l'applicazione software. La velocità di digitazione e i click del mouse sono i parametri principali per il rilevamento e il riconoscimento delle emozioni umane di base.

2.2.2 Approcci alla raccolta dei dati

I metodi di raccolta dei dati sono espliciti, impliciti e ibridi [28]. La raccolta dei dati può essere rispettivamente manuale, semiautomatica e automatica. L'accuratezza del profilo dell'utente dipende dalla quantità di dati generati attraverso l'interazione tra utente e sistema; inoltre, da dove vengono conservate le informazioni raccolte o elaborate; dal processo di personalizzazione in termini di quando e dove le informazioni sono disponibili per essere sfruttate per la personalizzazione.

2.2.2.1 Dati espliciti. Gli approcci di raccolta di informazioni esplicite sugli utenti sono acquisiti con tecniche manuali attraverso l'intervento degli utenti, come moduli di registrazione, questionari o siti di formazione classificati dagli utenti, o chiedendo agli utenti di valutare gli articoli, o tracciando le parole delle query degli utenti.

Ad esempio, eBay chiede agli utenti di fornire le loro opinioni e di dare un voto ai servizi e ai prodotti offerti. L'azienda utilizza queste informazioni per migliorare le raccomandazioni personalizzate fornite ai clienti. My Yahoo chiede esplicitamente all'utente di fornire informazioni personali che vengono memorizzate per creare un profilo.

Nei sistemi di raccomandazione, i dati di valutazione espliciti sono ampiamente utilizzati per profilare le preferenze degli utenti. Anche i tag sono comunemente utilizzati come parole chiave di interesse diretto quando vengono allegati dall'utente a contenuti web o utilizzando siti web sociali [14].

Alcuni ricercatori utilizzano direttamente i commenti e i post degli utenti per estrarre parole chiave che rappresentino gli interessi degli utenti [19], mentre altri utilizzano direttamente gli elementi valutati come indicazione degli interessi degli utenti [20].

I profili espliciti hanno inoltre una natura statica e sono validi solo finché l'utente non modifica esplicitamente i propri parametri di interesse e preferenza. Per superare questo problema, si estraggono le informazioni in modo implicito, tracciando le attività e/o i comportamenti degli utenti [28].

2.2.2.2 Dati impliciti. Il metodo implicito cerca di dedurre gli interessi o il contesto dell'utente dai log elaborati o dalle informazioni raccolte automaticamente senza alcuno sforzo da parte dell'utente. A differenza della profilazione statica, la profilazione automatica utilizza il metodo implicito e analizza il modello di comportamento dell'utente (ad esempio, la cronologia di utilizzo) per determinare gli interessi

dell'utente e sviluppare la sua conoscenza di base. La profilazione implicita dell'utente può essere sia lato client che lato server.

Gli agenti del browser, come quelli utilizzati in [15], raccolgono informazioni in modo interattivo, mentre l'utente naviga. Devono essere installati sul computer desktop dell'utente per catturare l'attività dell'utente e le azioni eseguite sulla pagina Web, come il tagging, il bookmarking e il download. Shen et al. [29] realizzano un feedback implicito su un agente di ricerca web intelligente lato client (UCAIR) basato sull'espansione della query e sulle informazioni di click-through.

Fernandez et al. [7] hanno previsto gli attributi demografici degli utenti del Web in base ai loro comportamenti di navigazione. Moawad et al. [21] hanno proposto un sistema basato su agenti per inizializzare i risultati della ricerca sul Web. Questo modello costruisce un profilo implicito dell'utente a partire dai segnalibri e dalle attività di navigazione utilizzando i weblog e le interazioni di ricerca utilizzando i log di ricerca. La query dell'utente viene ottimizzata semanticamente utilizzando gli interessi del profilo utente e WordNet.

Tuttavia, diversi progetti [17], [21], [28], [14] e [27] sono stati in grado di fornire ricerche personalizzate con successo costruendo profili utente basati su queste informazioni. Gli svantaggi di questo approccio sono la disponibilità di una quantità di dati molto inferiore e, a volte, i problemi di privacy per gli utenti.

2.2.2.3 Dati ibridi. I profili utente ibridi sono ottenuti con tecniche semiautomatiche e con un coinvolgimento limitato dell'utente. Il sistema raccoglie informazioni sull'utente e sull'utilizzo in un approccio misto di metodi impliciti, manuali ed espliciti. Questo approccio consente di ottenere profili più efficienti e di mantenere l'accuratezza delle informazioni temporali, poiché le informazioni vengono aggiornate temporalmente [18],[31].

Kobsa [16] ha creato una profilazione basata su parole chiave combinando feedback espliciti come forme di informazioni personali e un insieme di parole chiave di interesse e fonti di informazioni implicite come segnalibri web e cronologie di navigazione dell'utente, rispettivamente. Un altro sistema di profilazione ibrido [30] ha generato profili espliciti e impliciti in modo indipendente, per poi combinarli in un unico profilo. Il profilo utente esplicito si limita a chiedere agli utenti di specificare i loro interessi in termini di parole chiave. Il profilo utente implicito viene creato monitorando costantemente le attività dell'utente, memorizzando la cronologia di navigazione, il tempo trascorso su ogni pagina e le azioni aggiuntive come la stampa o il salvataggio. Combina gli interessi dichiarati esplicitamente con l'osservazione del comportamento dell'utente.

Fakhfakh [6] ha utilizzato modelli espliciti e impliciti per l'acquisizione delle preferenze degli utenti. In un modo esplicito, hanno intervistato 30 utenti per conoscere le loro preferenze e nell'altro modo, le preferenze degli utenti sono state scoperte attraverso i dataset di Movie Lens raccolti dal suo sito web che rappresenta un servizio di raccomandazione di film.

2.3 Costruzione e modellazione del profilo utente

2.3.1 Rappresentazione del profilo utente

Questa sotto-sezione riguarda la seconda fase della profilazione dell'utente, ovvero la memorizzazione e la rappresentazione delle informazioni raccolte nella fase precedente.

Questi profili, riassunti in [9], sono rappresentati da una raccolta di dati precedentemente preparati che riflettono gli interessi dell'utente. In questa sezione discutiamo tre tipi di modelli utente: basati su vettori, reti semantiche e concetti.

2.3.1.1 Basato sui Termini (Modello vettoriale-spaziale). Un modello utente basato su parole chiave o vettoriale è la rappresentazione più comune degli interessi degli utenti. È costituito da una raccolta di dati

denominata insieme di termini (o spazio vettoriale di termini) da vettori ponderati di parole chiave che possono essere utilizzati direttamente per espandere le query degli utenti [29].

Booleano, (TF) e (TF-IDF) sono metodi diversi per sviluppare una rappresentazione per ogni termine ponderato dei documenti o delle query. Inoltre, la somiglianza tra due vettori di termini ponderati viene calcolata mediante la somiglianza del coseno (CS). Le parole chiave degli interessi dell'utente vengono estratte dai documenti visitati durante la navigazione. Sono rappresentate da un singolo vettore che include tutti gli interessi o da vettori multipli, che riflettono l'interesse per diversi domini [9].

In questo modello, l'efficacia dei profili utente dipende dal grado di generalizzazione dei vettori. I suoi principali svantaggi sono la polisemia e il fatto che può causare una cattiva interpretazione degli interessi degli utenti. Yasser et al. [22] hanno migliorato il loro modello iniziale basato sui termini (IUM) costruendo un modello utente semanticamente migliorato (SUM) in cui i termini sono mappati a concetti di alto livello correlati.

Per classificare gli studenti in base ai loro interessi e al modello di Felder-Silverman viene utilizzata una tecnica fuzzy. Wang et al. [12] hanno proposto un nuovo metodo di algoritmo di estrazione basato sulla combinazione di Word2Vec, uno strumento di addestramento per vettori di parole pubblicato da Google nel 2013, e TF-IDF per rappresentare il vettore dei pesi, al fine di risolvere le carenze del metodo di rappresentazione vettoriale.

2.3.1.2 Basato sulla rete semantica (Ontologia). Per risolvere i problemi di ambiguità polisemica e sinonimia insiti nei profili basati su parole chiave, i profili possono essere rappresentati da un concetto ponderato di ciascun nodo della rete semantica. Tuttavia, questo comporta un onere per la facilità di costruzione di tale sistema. È necessaria una mappatura esistente tra le parole di interesse e i concetti, come ad esempio: l'ontologia WordNet, oppure attraverso la costruzione di un sistema di apprendimento o manualmente.

Liu [18] ha generato profili utente migliorati semanticamente che riflettono gli interessi e le intenzioni degli utenti in un insieme di categorie specifiche. In [31] è stato proposto un modello utente basato sull'ontologia, che facilita l'interazione semantica di un gran numero di applicazioni diverse con la stessa base di conoscenza e rappresenta gli interessi sotto forma di concetti ontologici interconnessi con l'ontologia di partenza predefinita, utilizzando i concetti di Wikipedia e WordNet.

Skillen et al. [32] hanno descritto il miglioramento del profilo ontologico dell'utente in ambienti mobili per un sistema context-aware.

In [34] è stato proposto un metodo per migliorare la presentazione nel social web utilizzando una rappresentazione ontologica della conoscenza che comprende FOAF, SIOC e DC e che consente di tracciare un profilo semantico nei siti web sociali.

2.3.1.3 Basato sui concetti (basato sulla gerarchia). I profili basati sui concetti sono identici a quelli basati sulle reti semantiche, in quanto rappresentano i nodi dei concetti e le relazioni tra questi nodi. Vengono applicati vari meccanismi per associare un peso a quanto l'utente è interessato a ciascun argomento. N. Nanas et al.

[23] hanno utilizzato diverse tecniche per categorizzare gli interessi degli utenti sotto forma di gerarchia e hanno utilizzato una base di conoscenza per questo scopo. I ricercatori hanno utilizzato il sistema di classificazione della biblioteca (LCS) per trovare le gerarchie di interessi degli utenti e il meccanismo (TF-IDF) per calcolare il peso dei bordi.

2.3.2 Costruzione del profilo utente

Quando si crea un profilo di una parola chiave è necessario considerare due aspetti. Il primo è quale fonte di informazioni utilizzare e il secondo è come creare il profilo con l'aiuto delle fonti selezionate. Le tecniche di creazione dei profili utente si basano sull'apprendimento automatico, sul data mining o sul recupero delle

informazioni. A seconda della rappresentazione del profilo utente desiderata, possono essere appropriate tecniche diverse. È importante che il metodo di costruzione scelto mantenga il profilo aggiornato per riflettere accuratamente le preferenze dell'utente; questo si è rivelato un compito molto impegnativo [14].

2.3.2.1 Recupero delle informazioni. I sistemi di recupero delle informazioni diventano sempre più interattivi e richiedono un apprendimento continuo con il sistema per migliorare il recupero. Dopo l'estrazione delle informazioni rilevanti, potrebbero esserci problemi di preelaborazione dei dati, come dati rumorosi, dati duplicati o dati che sembrano duplicati ma che potrebbero essere unici. Per risolvere questo problema, i dati unici e quelli duplicati sono necessari per effettuare la pulizia dei dati. Nell'approccio di *Information Retrieval*, sia il documento che gli interessi dell'utente sono rappresentati sotto forma di vettori di termini ponderati o presentando vettori di argomenti.

Questa strategia soffre di problemi di polisemia e sinonimia e, inoltre, del contenuto di pagine web non correlate come (homepage, contatti, ecc.).

Fakhfakh et al. [27] hanno sviluppato un framework basato su regole di associazione per la selezione del personale tra le caratteristiche del personale e le attività lavorative. Enrique et al. [8] hanno presentato diverse tecniche di data mining che possono essere utilizzate per catturare efficacemente il comportamento degli utenti.

2.3.2.2 Approccio di apprendimento automatico. L'apprendimento automatico è comunemente usato per la classificazione dei documenti in una raccolta per argomenti. Gli argomenti risultanti possono essere utilizzati come parole chiave o concetti per costituire un modello di utente. Alcuni lavori affrontano il problema della personalizzazione automatica del web utilizzando tecniche di apprendimento automatico per modellare il comportamento degli utenti e l'apprendimento degli interessi degli utenti. Gli alberi decisionali, il classificatore bayesiano ingenuo e i K Nearest Neighbors sono utilizzati per identificare gli argomenti dei documenti imparando le classificazioni da un insieme etichettato di documenti di addestramento.

Quickstep è un sistema di raccomandazione che ha classificato ogni documento di ricerca con un argomento, derivato dalle categorie informatiche di DMOZ, utilizzando l'apprendimento automatico [27]. Alcuni progetti hanno utilizzato il Latent Semantic Indexing (LSI) e il Linear Least Squares Fit (LLSF) per creare vettori di caratteristiche basati su parole chiave [36].

Song et al. [33] hanno utilizzato per prevedere alcuni dati sui compiti degli utenti un insieme di parametri latenti e la mappatura dei documenti con i termini.

Liu [18] ha modellato una preferenza dinamica dell'utente per la raccomandazione, utilizzando un modello topic probabilistico Latent Dirichlet Allocation (PLDA) per rappresentare la preferenza temporale dell'utente.

2.3.2.3 Approccio basato sulle ontologie. Gli approcci più recenti manifestano l'unione del web adattivo e del web semantico [34] utilizzando linguaggi di ontologia web standardizzati e la loro base software. In questo approccio, i profili degli utenti sono generati sulla base di un'ontologia di riferimento, ad esempio la gerarchia di Open Directory come ontologia per trovare gli interessi degli utenti [38].

L'ontologia di riferimento è in grado di risolvere o almeno ridurre il problema della specificità dell'applicazione dell'ontologia e potrebbe fornire importanti vantaggi rispetto all'ontologia di dominio e di applicazione precedentemente utilizzata [41].

L'uso dell'ontologia nell'acquisizione di domini di conoscenza sulle preferenze di un utente, pertanto questa ontologia è stata costruita sulla base della tassonomia standard degli annunci per i profili utente, della tassonomia OpenDNS, dell'ontologia FOAF e dell'ontologia DBpedia. Il profilo utente ontologico [18] consente di identificare e ragionare facilmente sugli argomenti di interesse di un utente.

Un profilo utente è dimostrato in formato RDF (Resource Description Framework) / XML o OWL (Ontology Web Language). Il profilo dell'utente è stato costruito automaticamente estraendo gli URL postati dagli utenti nei tweet, l'ontologia del profilo dell'utente, OpenDNS, DBpedia e i database NoSQL e Reasoner.

Bouzeghoub [5] ha proposto un modello di profilo dell'allievo multi-agente che unisce gli standard IMS LIP e FOAF alle informazioni sul contesto e sull'agenda, raccolte dai fornitori di social network.

2.3.2.4 Approccio basato su agenti. La tecnica del sistema ad agenti è utilizzata per migliorare i criteri di richiamo e di precisione del reperimento delle informazioni attraverso la costruzione di agenti multipli per affrontare le diverse sfide e fasi della personalizzazione. Consente di creare esperimenti sotto forma di modelli e di eseguirli per fornire come output la rappresentazione dell'interazione tra gli agenti. In [10] è stato proposto un modello di ricerca web personalizzata (WSP) basato su agenti per recuperare risultati di ricerca migliori per le preferenze dell'utente basati sulla tecnica del sistema multi-agente, oltre a risolvere il problema della polisemia e della sinonimia per il profilo basato su parole chiave utilizzando l'ontologia Word Net e sfruttando le preferenze dell'utente. I ricercatori hanno proposto altri metodi, come il modello basato su agenti per le raccomandazioni di notizie personalizzate e consapevoli del contesto.

2.4 Personalizzazione

La terza fase prevede lo sfruttamento delle informazioni contenute nel profilo dell'utente per fornire servizi personalizzati. Una volta costruito il profilo dell'utente, questo viene utilizzato per fornire servizi personalizzati in diversi ambiti, come i sistemi di raccomandazione personalizzati, i servizi mobili, i social network, la ricerca e la navigazione sul Web, ecc.

2.4.1 Identificazione dell'utente

L'identificazione accurata dell'utente è essenziale per qualsiasi sistema che costruisca profili che rappresentino i singoli utenti. Esistono cinque approcci di base all'identificazione dell'utente: agenti software, login, server proxy avanzati, cookie e id di sessione.

Qian Gao et al. [15], utilizzano due metodi di opzione adattati per identificare gli utenti: uno è quello dei cookie per le sessioni correnti e l'altro è quello dei login per gli utenti che scelgono di registrarsi in un sito. Il web data mining può essere utilizzato per identificare gli utenti e trovare le preferenze degli studenti sulla base dei weblog esistenti sui server [22].

I weblog registrano ogni transazione eseguita dal browser a ogni accesso al web. Grazie a questi dati, possiamo seguire le sequenze di visita delle singole pagine da parte dell'allievo che, a una certa condizione, è identificato dall'indirizzo IP.

2.4.2 Applicazioni

2.4.2.1 Ricerca e navigazione web personalizzata. La ricerca personalizzata è il processo di ricerca di un insieme di documenti o pagine Web privilegiando gli interessi dell'utente. L'interesse dell'utente viene identificato dal profilo dell'utente, che viene creato analizzando i siti ricercati dall'utente. L'accuratezza del sistema è migliorata grazie alla previsione dell'interesse dell'utente rispetto a una ricerca normale.

In [3] è stata proposta una rete concettuale per recuperare le pagine web in base alle preferenze dell'utente e migliorare la navigazione tenendo conto del dinamismo delle preferenze e dei nuovi concetti durante la ricerca.

Una recente ricerca segue la profilazione basata sull'ontologia per approcci di ricerca personalizzati [28] utilizzando la tassonomia dell'Open Directory Project (ODP) come ontologia degli argomenti del Web.

2.4.2.2 Sistemi di raccomandazione e previsione. Un sistema di raccomandazione è una sottoclasse del sistema di filtraggio delle informazioni che cerca di prevedere la valutazione degli articoli o la preferenza dell'utente. Un sistema di raccomandazione personalizzato ha due aspetti importanti: conoscere l'utente e, in base a ciò, raccomandare gli articoli di suo interesse.

Thorat et al. [37] hanno analizzato diverse tecniche di sistema di raccomandazione basate sulla demografia, sul filtraggio basato sui contenuti, sul filtraggio collaborativo e sul filtraggio ibrido. La profilazione dell'utente è in grado di ottenere informazioni adeguate e accurate sugli interessi e le caratteristiche di un utente e di stabilirle con un intervento minimo da parte dell'utente [2] per offrire una consapevolezza appropriata con una raccomandazione di mitigazione pertinente basata sulla posizione di sicurezza.

Hong et al. [11] hanno proposto un sistema di raccomandazione del lavoro basato sul tempo e sulla dimensionalità. In questo caso, hanno aggiornato dinamicamente il profilo del candidato sulla base degli insiemi di caratteristiche degli interessi dell'utente, estratti in un periodo di tempo, attraverso il concetto di guadagno di informazioni.

Zhang et al. [42] hanno esplorato un profilo basato sulla gerarchia, creando tassonomie basate sui termini della descrizione degli articoli. In [1] è stata sviluppata una tecnica interessante per risolvere il problema della rappresentazione del profilo utente ad albero di concetti. Viene sviluppato un modello Dynamic Normalized Tree of Concepts (DNTC) per i profili degli utenti e un sistema di raccomandazione di articoli di ricerca utilizzando la ntologia ACM 2012.

Quercia et al. [26] hanno studiato la correlazione tra la personalità degli utenti e gli attributi dei loro profili social web. Hanno indicato che alcuni tratti della personalità, come l'estroversione e il 'nevrotismo', possono essere accuratamente previsti in base al profilo dell'utente su Twitter. È stato proposto un approccio [18] basato sulla generazione di profili degli utenti, al fine di fornire pubblicità e servizi che corrispondono ai loro interessi, estratti e ottenuti attraverso gli URL che condividono i loro tweet.

2.4.2.3 Sistemi di apprendimento adattivo. L'apprendimento adattivo è una branca dell'e-learning che aiuta gli studenti nel loro percorso di apprendimento e di studio. Nguyen [24] ha proposto un framework di modellazione dell'utente chiamato Zebra, che gestisce il Modello Triangolare dell'Apprendente (TLM) delle caratteristiche degli studenti e fornisce un potente meccanismo di inferenza per ragionare su nuove informazioni sugli studenti.

2.4.2.4 Visualizzazione. Yingze et al. [39] hanno riportato uno studio di usabilità dell'interfaccia *Introspective-Views* per la visualizzazione di modelli semantici di utenti nelle reti sociali. Hanno descritto come l'interfaccia possa essere utilizzata in una rete sociale per visualizzare e modificare i modelli di interesse degli utenti basati sull'ontologia.

2.4.2.5 Servizi personalizzati di social network online. Le applicazioni di social network personalizzate classificano automaticamente i dati in categorie diverse ma più rilevanti e forniscono agli utenti i dati più pertinenti ai loro interessi. È stata presentata una tecnica di profilazione delle preferenze dell'utente [14] che genera un modello di classificazione per le preferenze dell'utente basato sul comportamento dell'utente su Facebook, come like, share e post. Il modello viene creato a partire da tre modelli di classificazione e conclude che la Support Vector Machine ha prestazioni migliori rispetto agli algoritmi ANN e NB. Una tecnica proposta utilizza Folksonomies per raccogliere i dati personali e le personalità degli utenti, attraverso un sistema di tagging collaborativo [40].

2.5 Analisi critica

Nel corso dello studio sulla profilazione degli utenti, sono state individuate diverse lacune nella ricerca e nella letteratura precedente sugli approcci alla profilazione degli utenti e alla personalizzazione.

Queste lacune includono

- Necessità di migliorare l'accuratezza del recupero della percezione dell'utente con le caratteristiche fisiche dei dati. Pertanto, dovrebbe esistere un approccio di apprendimento per verificare l'input dell'utente utilizzando un feedback di rilevanza [29].

- Con la potenza dell'information retrieval e dell'apprendimento automatico, si deve scoprire implicitamente il profilo ontologico dell'utente e affrontare il problema dell'avvio a freddo e della scarsità dei profili utente [25].
- Vi è la necessità di estrarre l'incertezza insita nei dati del profilo associando ai profili dei pesi, attraverso tecniche di ragionamento esplicito [9].
- Vi è bisogno di un modello basato su un'ontologia per descrivere la varietà di elementi di interesse e di profili utente. Per aumentare l'accettazione dei modelli basati sulla fiducia dei sistemi di raccomandazione semantici [20].
- C'è bisogno di una maggiore considerazione dei modelli di fiducia basati su topic/tag nel contesto di contenuti eterogenei per l'elaborazione del web [19]. È stata prestata poca attenzione allo sfruttamento delle strutture a grafo e a link tra risorse e persone, soprattutto nel contesto dei social media online [40].

Bibliografia di riferimento

- [1] Modhi Al Alshaikh, Gulden Uchyigit, and Roger Evans. 2017. A research paper recommender system using a Dynamic Normalized Tree of Concepts model for user modelling. In Research Challenges in Information Science (RCIS), 2017 11th International Conference on. IEEE, 2005210.
- [2] Bhushan S Atote, Taranpreet Singh Saini, Mangesh Bedekar, and Saniya Zahoor. 2016. Inferring emotional state of a user by user profiling. In Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on. IEEE, 530535.
- [3] KR Remesh Babu and Philip Samuel. 2015. Concept Networks for Personalized Web Search Using Genetic Algorithm. *Procedia Computer Science* 46 (2015), 566573.
- [4] Nor Adzlan bin Jamaludin, Muthukkaruppan Annamalai, Nursuriati Jamil, and Zainab Abu Bakar. 2013. A model for keyword Profile Creation using extracted keywords and terminological ontology. In e-Learning, e-Management and e-Services (IC3e), 2013 IEEE Conference on. IEEE, 1365141.
- [5] Amel Bouzeghoub and Alda Lopes Gancarski. 2011. An agent-based service architecture for user profiles dynamic share. In 2011 Seventh International Conference on Semantics, Knowledge and Grids. IEEE, 1375143.
- [6] Rim Fakhfakh, Ghada Feki, Anis Ben Ammar, and Chokri Ben Amar. 2016. Personalizing information retrieval: a new model for user preferences elicitation. In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 0020915002096.
- [7] Miriam Fernandez, Arno Scharl, Kalina Bontcheva, and Harith Alani. 2014. User profile modelling in online communities. (2014).
- [8] Enrique Frias-Martinez, Sherry Y Chen, and Xiaohui Liu. 2006. Survey of data mining approaches to user modeling for adaptive hypermedia. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 36, 6 (2006), 7345749.
- [9] Susan Gauch, Mirco Speretta, Aravind Chandramouli, and Alessandro Micarelli. 2007. User profiles for personalized information access. In *The adaptive web*. Springer, 54589.
- [10] Ahmad Hawalah and Maria Fasli. 2011. A multi-agent system using ontological user profiles for dynamic user modelling. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*. IEEE Computer Society, 4305437.
- [11] Wenxing Hong, Siting Zheng, and Huan Wang. 2013. Dynamic user profile-based job recommender system. In *Computer Science & Education (ICCSE)*, 2013 8th International Conference on. IEEE, 149951503.
- [12] Jianqiao Hu, Feng Jin, Guigang Zhang, Jian Wang, and Yi Yang. 2017. A User Profile Modeling Method Based on Word2Vec. In *Software Quality, Reliability and Security Companion (QRS-C)*, 2017 IEEE International Conference on. IEEE, 4105414.
- [13] Rachsuda Jiamthapthaksin and Than Htike Aung. 2017. User preferences profiling based on user behaviors on Facebook page categories. In *Knowledge and Smart Technology (KST)*, 2017 9th International Conference on. IEEE, 2485253.

- [14] Diane Kelly and Jaime Teevan. 2003. Implicit feedback for inferring user preference: a bibliography. In *Acm Sigir Forum*, Vol. 37. ACM, 18528.
- [15] JeeHyun Kim, Qian Gao, and Young Im Cho. [n. d.]. A Context-Aware based Dynamic User Preference Profile Construction Method. ([n. d.]).
- [16] Alfred Kobsa. 2001. Generic user modeling systems. *User modeling and user adapted interaction* 11, 1-2 (2001), 49563.
- [17] Lin Li, Zhenglu Yang, Botao Wang, and Masaru Kitsuregawa. 2007. Dynamic adaptation strategies for long-term and short-term user profile to personalize search. In *Advances in Data and Web Management*. Springer, 2285240.
- [18] Xin Liu. 2015. Modeling Users' Dynamic Preference for Personalized Recommendation. In *IJCAI*. 178551791.
- [19] Chunliang Lu, Wai Lam, and Yingxiao Zhang. 2012. Twitter user modeling and tweets recommendation based on wikipedia concept graph. In *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*. 33538.
- [20] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2875296.
- [21] Ibrahim F Moawad, Hanaa Talha, Ehab Hosny, and Mohamed Hashim. 2012. Agent-based web search personalization approach using dynamic user profile. *Egyptian Informatics Journal* 13, 3 (2012), 1915198.
- [22] Yasser A Nada and Khaled M Fouad. 2011. An approach to improve the representation of the user model in the web-based systems. *IJACSA) International Journal of Advanced Computer Science and Applications* 2, 12 (2011), 1525160.
- [23] Nikolaos Nanas, Victoria Uren, and Anne De Roeck. 2003. Building and applying a concept hierarchy representation of a user profile. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 1985204.
- [24] Loc Nguyen. 2014. A user modeling system for adaptive learning. In *Interactive Collaborative Learning (ICL), 2014 International Conference on*. IEEE, 8645866.
- [25] Michael J Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In *The adaptive web*. Springer, 3255341.
- [26] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our twitter profiles, our selves: Predicting personality with twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 1805185.
- [27] Krishnan Ramanathan, Julien Giraudi, and Ajay Gupta. 2008. Creating hierarchical user profiles using Wikipedia. *HP Labs* 6 (2008), 1635170.
- [28] Silvia Schiaffino and Analía Amandi. 2009. Intelligent user profiling. In *Artificial Intelligence An International Perspective*. Springer, 1935216.
- [29] Xuehua Shen, Bin Tan, and ChengXiang Zhai. 2005. Implicit user modeling for personalized search. In *Proceedings of the 14th ACM international conference on Information and knowledge management*. ACM, 8245831.
- [30] P. Shitole and M. Potey. 2015. Focusing User Modeling For Age Specific Differences. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* 4, 3 (2015).
- [31] Ahu Sieg, Bamshad Mobasher, and Robin Burke. 2007. Web search personalization with ontological user profiles. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, 5255534.
- [32] Kerry-Louise Skillen, Liming Chen, Chris D Nugent, Mark P Donnelly, William Burns, and Ivar Solheim. 2012. Ontological user profile modeling for context aware application personalization. In *International Conference on Ubiquitous Computing and Ambient Intelligence*. Springer, 2615268.
- [33] Rui Song, Enhong Chen, and Min Zhao. 2005. SVM based automatic user profile construction for personalized search. In *International Conference on Intelligent Computing*. Springer, 4755484.
- [34] Sergey Sosnovsky and Darina Dicheva. 2010. Ontological technologies for user modelling. *International Journal of Metadata, Semantics and Ontologies* 5, 1 (2010), 32571.
- [35] Hamza Hydri Syed and Periklis Andritsos. 2007. User preference modeling-a survey report. Technical Report. University of Trento.
- [36] Jie Tang, Limin Yao, Duo Zhang, and Jing Zhang. 2010. A combination approach to web user profiling. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 5, 1 (2010), 2.

- [37] Poonam B Thorat, RM Goudar, and Sunita Barve. 2015. Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications* 110, 4 (2015).
- [38] Joana Trajkova and Susan Gauch. 2004. Improving ontology-based user profiles. In *Coupling approaches, coupling media and coupling languages for information retrieval*. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, 380\$390.
- [39] Yingze Wang and Shi-Kuo Chang. 2014. User Profile Visualization to Facilitate MSLIM-Model-Based Social Influence Analysis Based on Slow Intelligence Approach. *International Journal of Software Engineering and Knowledge Engineering* 24, 10 (2014), 1573\$1587.
- [40] Au Yeung, Ching Man, Nicholas Gibbins, and Nigel Shadbolt. 2008. A study of user profile generation from folksonomies. (2008).
- [41] Leila Zemmouchi-Ghomari and Abdessamed Reda Ghomari. 2009. Reference Ontology. In *2009 Fifth International Conference on Signal Image Technology and Internet Based Systems*. IEEE, 485\$491.
<https://doi.org/10.1109/SITIS.2009.81>
- [42] Yuchen Zhang, Amr Ahmed, Vanja Josifovski, and Alexander Smola. 2014. Taxonomy discovery for personalized recommendation. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 243\$252.
- [43] Marina Farid, Ibrahim Moawad, Rania Elgohary, Mohamed Roushdy 2018, User Profiling Approaches, Modeling, and Personalization. In *INFOS'18*, December 2018, CAIRO, EGYPT