

# ACCESSO PERSONALIZZATO A DOCUMENTI MULTIVERSIONE PER APPLICAZIONI NEL SETTORE DELL'E-GOVERNMENT \*

Fabio Grandi, Maria Rita Scalas  
DEIS, Alma mater Studiorum – Università di Bologna  
Viale Risorgimento 2, I-40136, Bologna, Italy  
E-mail: {fgrandi,mrscalas}@deis.unibo.it

Federica Mandreoli, Riccardo Martoglia, Enrico Ronchetti, Paolo Tiberio  
DII, Università di Modena e Reggio Emilia  
Via Vignolese 905/b, I-4110, Modena, Italy  
E-mail: {fmandreoli,rmartoglia,eronchetti,ptiberio}@unimo.it

**ABSTRACT.** *In questo lavoro viene presentata l'attività di ricerca concernente la realizzazione di sistemi prototipali per la gestione efficiente di documenti XML multiversione in uno scenario di e-Government. Lo scopo applicativo di tali sistemi è di permettere al cittadino l'accesso a versioni personalizzate di risorse quali testi normativi e informazioni rese disponibili sul WEB dalle Pubbliche Amministrazioni. Per rappresentare l'evoluzione delle norme nel tempo e il conseguente "versionamento" si sono usate quattro dimensioni temporali e un'ulteriore dimensione semantica per rappresentare l'applicabilità delle norme a differenti classi di cittadini, in accordo alla loro identità digitale. La classificazione dei cittadini è basata sulla gestione di un'ontologia e l'adozione di tecniche di Semantic WEB. L'attuale implementazione, evoluzione di un approccio di tipo "stratum" (sviluppato on top di una piattaforma RDBMS), è basata su un approccio "nativo" consistente in un query processor XML sviluppato ad-hoc. Una sperimentazione preliminare ha evidenziato nel nuovo sistema buoni livelli di prestazioni e scalabilità.*

**KEYWORDS:** e-Government, XML, basi di dati temporali, Semantic Web.

## 1. INTRODUZIONE

In questo lavoro presentiamo le nostre attività di ricerca concernenti la realizzazione di Sistemi Informativi WEB per applicazioni nel campo dell'e-Government [6, 8]. Più precisamente, il nostro approccio fa uso di tecniche sviluppate per le basi di dati temporali e per il *semantic WEB* per fornire un accesso **personalizzato** a risorse e servizi multiversione messi a disposizione dei cittadini da una Pubblica Amministrazione. L'offerta di versioni personalizzate ha lo scopo di incrementare e ottimizzare il coinvolgimento dei cittadini nel processo di e-Governance. In particolare, come tipo di servizio consideriamo l'accesso selettivo a testi normativi e a documenti resi disponibili sul Web in formato XML.

Innanzitutto l'elevato grado di dinamismo insito nei sistemi normativi ha come conseguenza la coesistenza di versioni multiple dei testi normativi, dato che le leggi sono continuamente soggette a emendamenti e modifiche. Infatti è di importanza cruciale riuscire a ricostruire la **versione consolidata** di una norma così come ottenuta dall'applicazione di tutte le successive modifiche subite fino ad oggi, che è la forma in cui attualmente fa parte dell'ordinamento e nella quale deve essere applicata oggi. In ogni caso, anche le versioni passate sono importanti, non solo per ragioni storiche ma anche, per esempio, per permettere ad un tribunale di pronunciarsi oggi su fatti commessi nel passato. In altre parole, riferimenti temporali sono onnipresenti nel dominio dell'e-Government e un sistema informativo legale dovrebbe essere in grado di reperire o ricostruire a richiesta qualsiasi versione di un dato documento per soddisfare un comune

---

\* Questo lavoro è stato finanziato dal progetto MIUR-PRIN "Il cittadino Europeo nell'e-Governance: profili filosofico-giuridici, giuridici, informatici ed economici" (<http://www.cirsfid.unibo.it/eGov03/>).

requisito applicativo. In tale contesto, pertanto, la personalizzazione è basata sulla prospettiva temporale dell'utente.

Inoltre, in uno scenario di e-Government, c'è un altro tipo di versionamento che gioca un ruolo importante perché alcuni documenti, o alcune loro parti, hanno o acquisiscono una limitata applicabilità. Per esempio una data norma (che per esempio definisce il regime fiscale) può contenere alcuni articoli che sono applicabili a differenti classi di cittadini: un articolo è applicabile esclusivamente alle persone disoccupate, un articolo ai lavoratori autonomi, e così via. Quindi un cittadino che utilizza un sistema di accesso ad una raccolta normativa può essere interessato a reperire una versione "su misura" delle norme di interesse, cioè una versione che contenga soltanto gli articoli che siano applicabili al suo caso personale. Quindi, la personalizzazione in questo contesto è basata sulla applicabilità limitata al caso del cittadino ed è necessario un **versionamento semantico** dei documenti memorizzati.

In questo contesto abbiamo definito modelli di dati per documenti XML multiversione e abbiamo costruito sistemi prototipali per la loro gestione efficiente in uno scenario applicativo di e-Government basato sul Web, che include un accesso *online* personalizzato a collezioni di norme. In particolare, in questo lavoro descriveremo l'attuale sistema, prodotto di una ricerca tuttora in corso che è stata presentata in [2] e che rappresenta l'evoluzione del primo sistema sviluppato secondo un approccio di tipo "*stratum*", cioè "*on top*" ad un DBMS commerciale (una descrizione e una valutazione più dettagliate sono state pubblicate in precedenza in [1,3]). Il prototipo è stato implementato utilizzando un approccio "nativo" ed è attualmente in fase di sperimentazione.

## 2. VERSIONAMENTO TEMPORALE E SEMANTICO

In una prima fase della nostra ricerca ci siamo concentrati sugli aspetti temporali e sulla gestione efficiente di testi normativi che variano nel tempo. A questo scopo abbiamo sviluppato un modello temporale di dati XML che usa quattro dimensioni temporali per rappresentare correttamente l'evoluzione delle norme nel tempo e il loro versionamento risultante. Le dimensioni considerate sono: Tempo di Validità, Tempo di Efficacia, Tempo di Transazione e Tempo di Pubblicazione [1]. Il modello è stato definito tramite uno Schema XML [10], dove la struttura delle norme è definita per mezzo di una gerarchia articolato-sezione-articolo-comma, e versioni multiple di contenuto possono essere definite ad ogni livello della gerarchia. Ogni versione è caratterizzata da attributi di tipo *timestamp*, che definiscono la sua pertinenza temporale rispetto a ciascuna delle dimensioni temporali di validità, di efficacia e di transazione.

In una seconda fase della ricerca, il modello multiversione basato sulle dimensioni temporali è stato esteso per includere un'ulteriore dimensione semantica per supportare l'accesso personalizzato ai testi normativi. In generale, la comprensione automatica delle informazioni disponibili sul Semantic Web richiede la marcatura semantica dei contenuti e la disponibilità di tool di ragionamento automatico. Per fare in modo che le informazioni e la loro interpretazione sia condivisa da più agenti, inclusi i tool automatici, si rende necessaria l'introduzione di ontologie di riferimento condivise [7]. Nel nostro caso abbiamo definito una ontologia civica, che corrisponde a una classificazione dei cittadini basata sulle distinzioni introdotte da norme (norme fondanti) che susseguendosi comportano limitazioni, totali o parziali, nella loro applicabilità. Pertanto nel nostro modello esteso la nuova dimensione di versionamento codifica informazioni circa l'applicabilità delle differenti parti di un testo normativo alle classi pertinenti all'interno dell'ontologia civica. Per maggiori informazioni sulla gestione dell'ontologia e delle relative marcature si veda [4].

Le raccolte di testi normativi online sono generalmente gestite tramite sistemi tradizionali di Information Retrieval, dove gli utenti possono accedere ai contenuti per mezzo di interrogazioni basate su parole chiave che caratterizzano l'argomento di loro interesse. Noi abbiamo esteso tale contesto offrendo all'utente la possibilità di esprimere interrogazioni contenenti quattro tipi di vincoli: temporali, strutturali, testuali e di applicabilità. Tali tipi di vincoli sono completamente ortogonali e permettono all'utente di effettuare ricerche molto accurate nella *repository* XML

contenente le norme. Poniamo che Mario Rossi sia un lavoratore autonomo interessato a tutte le norme che:

- contengono paragrafi (*vincolo strutturale*) che trattano di sanità (*vincolo testuale*);
- che erano valide ed efficaci tra il 2002 e il 2004 (*vincoli temporali*);
- che sono applicabili al suo caso personale (*vincolo di applicabilità*).

Tale *query* può essere formulata nel nostro sistema usando la sintassi standard XQuery come mostrato nel seguito:

```
FOR $a IN path
  WHERE textConstr ($a//paragraph//text(), 'health AND care')
    AND tempConstr ('vTime OVERLAPS PERIOD(2002-01-01,2004-12-31)')
    AND tempConstr ('eTime OVERLAPS PERIOD(2002-01-01,2004-12-31)')
    AND applConstr ('self-employed')
  RETURN $a
```

dove *textConstr*, *tempConstr* e *applConstr* sono funzioni dedicate rispettivamente alla specifica dei vincoli testuali, temporali, e di applicabilità (il vincolo strutturale è implicito nella espressione XPath usata nell'istruzione XQuery). Per mezzo dei vincoli temporali di validità ed efficacia, il cittadino è in grado di estrarre dal *repository* multiversione versioni correnti consolidate, o accedere a versioni passate di particolari testi normativi, sempre ricostruite in modo consistente dal sistema, sulla base delle necessità dell'utente e personalizzate secondo la sua identità digitale. L'identità digitale del cittadino è definita come insieme di tutte le informazioni che lo/la riguardano, e che sono necessarie ai fini della classificazione rispetto all'ontologia [2,9]. Tali informazioni devono essere reperibili in modo automatico nelle basi di dati gestite dalle Pubbliche Amministrazioni. A tale scopo, funzioni per l'interrogazione di tali basi di dati dovranno essere predisposte e implementate tramite servizi standard di accesso. Il confronto tra l'identità del cittadino e le classi dell'ontologia dovrà venire poi effettuato tramite un servizio di *reasoning* integrato nel sistema di gestione.

### 3. IMPLEMENTAZIONE E CONCLUSIONI

Tutti gli aspetti di interrogazione multidimensionale e di accesso personalizzato a norme in formato XML sono stati implementati nel sistema ad approccio "XML nativo", evoluzione del precedente sistema "stratum". L'architettura è mostrata nella parte destra di Fig. 1 ed è composta di un modulo **Temporal XML Query Processor** in grado di gestire la *repository* dei dati XML e di elaborare in un singolo componente tutti e quattro i tipi di vincoli rilevanti per le interrogazioni: temporali, strutturali, testuali e di applicabilità. Il prototipo è implementato su piattaforma Java JDK 1.5 e sfrutta strutture dati *ad hoc* (appoggiandosi su librerie DBMS integrate "leggere") e algoritmi che permettono all'utente di memorizzare e ricostruire al volo testi normativi che soddisfano i quattro tipi di vincoli. Differentemente dall'approccio *stratum* che abbiamo usato nel precedente prototipo (parte sinistra di Fig. 1), dove i vincoli temporali erano processati separatamente, tutti i vincoli strutturali, testuali e temporali sono trattati insieme dal **Temporal XML Query Processor**. Tale componente memorizza le norme XML non sotto forma di documenti interi ma convertendole in una collezione di ennuple temporali, ciascuna delle quali rappresenta una delle sue parti multiversione (cioè articoli, commi e così via). Queste strutture dati sono poi usate per eseguire in modo efficiente algoritmi di tipo *join strutturale* [5] che sono stati ideati e messi a punto per il contesto multiversione temporale/semantico. I vincoli testuali (ricerca per *keywords*) sono gestiti per mezzo di *inverted index* come nell'approccio *stratum*. L'approccio nativo è in grado di accedere e recuperare i soli dati strettamente necessari, evitando di prelevare documenti XML interi e costruire strutture che abbisognano di grandi quantità di memoria (es. alberi DOM).

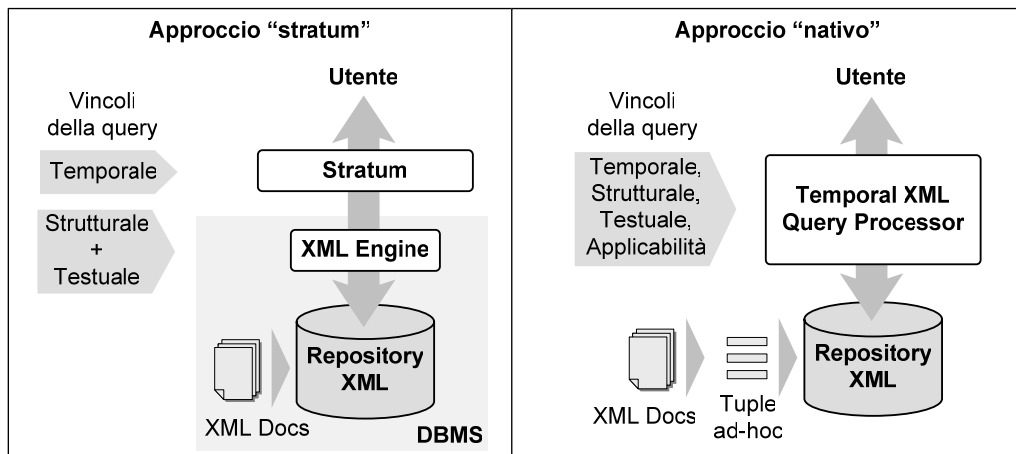


Fig. 1. Architetture del primo ("stratum") e del secondo (nativo) sistema a confronto

Un lavoro preliminare di sperimentazione sulle prestazioni, condotto su collezioni di documenti normativi XML sintetici, ha mostrato risultati incoraggianti [4]. L'efficacia dell'approccio nativo è stata dimostrata confrontando le sue prestazioni con quelle della nostra precedente implementazione di tipo *stratum*, introducendo un *benchmark* comune di interrogazioni e conducendo un certo numero di esperimenti per analizzare il suo comportamento nell'effettuare accessi personalizzati in base ai vincoli di applicabilità. In particolare l'approccio nativo ha mostrato la sua efficienza in un grande numero di situazioni sperimentali ed ha evidenziato eccellenti doti di scalabilità al variare della configurazione del carico di lavoro.

Il lavoro futuro prenderà in considerazione eventuali miglioramenti dell'approccio per fronteggiare requisiti applicativi più avanzati e il necessario completamento dell'infrastruttura con l'implementazione di servizi ausiliari (es. quello per la classificazione automatica rispetto all'ontologia dei cittadini connessi al servizio). Maggiori dettagli sul completamento dell'infrastruttura completa si possono trovare in [2,9]. Lavoro ulteriore sarà infine devoluto ad un *assessment* del sistema da noi sviluppato in un ambiente operativo concreto, con veri utenti e in presenza di una collezione di testi normativi reali.

## BIBLIOGRAFIA

- [1] Grandi, F. et al, 2003. A Temporal Data Model and Management System for Normative Texts in XML Format. *Proc. of the 15th ACM WIDM*, New Orleans, LA, pp. 29–36.
- [2] Grandi, F. et al, 2004. Management of the Citizen's Digital Identity and Access to Multi-version Norm Texts on the Semantic Web. *Proc. of IPSI 2004*, Pescara, Italy.
- [3] Grandi, F. et al, 2005a. Temporal Modelling and Management of Normative Documents in XML Format. *Data & Knowledge Engineering*, Vol. 54, No. 3, pp. 327–354.
- [4] Grandi, F. et al, 2005b. Personalized Access to Multi-version Norm Texts in an eGovernment Scenario. *Proc. of EGOV 2005 (in conj. with DEXA05)*, Copenhagen, Denmark, LNCS 3591, pp. 281–290.
- [5] Al-Khalifa, S. et al, 2002. Structural Joins: A Primitive for Efficient XML Query Pattern Matching. *Proc. of 18th ICDE*, San Jose, CA, pp. 141–154.
- [6] EC E-Gov, 2005. European Commission e-Government Home Page. [http://europa.eu.int/information\\_society/eeurope/2005/all\\_about/egovernment/index\\_en.htm](http://europa.eu.int/information_society/eeurope/2005/all_about/egovernment/index_en.htm)
- [7] Guarino, N., editor, 1998. *Formal Ontology in Information Systems*. IOS Press, Amsterdam, The Netherlands.
- [8] IT E-Gov, 2005. Il portale Italiano per l'e-Government: <http://www.italia.gov.it>
- [9] Mandreoli, F. et al, 2005. Efficient Management of Multi-version XML Documents for E-Government Applications. *Proc. of WEBIST 2005*, Miami, FL.
- [10] XMLSchema, 2004. The XML Schema Home Page. W3C Consortium, <http://www.w3c.org/XML/Schema/>