

Quaderni del DAE

Rivista di Diritto Amministrativo Elettronico

...

Regole di redazione e strumenti di *natural language processing*

per gli atti amministrativi

Pietro Mercatali e Francesco Romano

...

2005

1. Introduzione

La comprensione di un testo in linguaggio naturale avviene mettendo in relazione le parole che lo compongono e confrontando le informazioni così ottenute con le conoscenze possedute dal lettore o che egli può acquisire.

Ad esempio in un atto amministrativo non è scritto esplicitamente che quello è un atto amministrativo, ma le nostre conoscenze giuridiche, abbinata al succedersi delle parole, ci consentono di identificarlo come tale.

Il computer, al contrario dell'uomo, non è in grado di attribuire un significato a una sequenza di parole in base a conoscenze già acquisite, anche se è in grado di riconoscere una sequenza di parole e le relazioni che le legano.

Si definisce dunque il computer come macchina sintattica in quanto capisce i segni e le relazioni tra di essi ma non il loro significato.

La necessità di trasferire conoscenza semantica al computer ha indotto a introdurre nei testi dei codici o marcatori, tramite i quali si associano a determinate parole, o a loro combinazioni, dei significati, che in tal modo risultano leggibili e comprensibili anche dalla macchina.

Per tale operazione di trasferimento di conoscenza semantica si sta affermando, come standard, il linguaggio XML.

La marcatura di un testo con il linguaggio XML risulta comprensibile dalla macchina per mezzo di una apposita grammatica (o insieme di regole) che associa ad ogni marcatore un determinato significato.

La definizione di tale grammatica richiede la costruzione di *un modello di riferimento semantico-strutturale (DTD)* della classe di documenti che si vuol marcare.

La marcatura XML è attività che, se fatta manualmente, risulta lunga e dispendiosa.

Inoltre l'ambiguità e l'indeterminatezza del linguaggio naturale fa sì che i singoli documenti si distanzino spesso dal modello definito, aggravando ulteriormente il lavoro di marcatura.

Questa peculiarità del linguaggio espone il metodo di marcatura ad altri due pericoli:

- non riuscire a cogliere tale ambiguità ed indeterminatezza,
- piegare il significato testuale all'esigenza di utilizzo di tale marcatura.

Per tentare di ovviare a questi problemi si può agire in una duplice direzione:

1. sviluppare strumenti di NLP che consentano il riconoscimento del testo e la sua marcatura automatica anche se le strutture testuali si distanziano dal modello;
2. stabilire degli standard linguistico documentari (regole di scrittura dei documenti) che rendano più omogenee, coerenti e puntuali le strutture dei documenti e quindi consentano alla macchina di riconoscere più agevolmente le strutture standardizzate.

La definizione e l'introduzione di standard per la redazione dei documenti amministrativi non è però solo funzionale al loro trattamento automatico, ma sembra rispondere ad almeno altre tre esigenze avvertite, in particolare, da chi propone una modernizzazione della pubblica amministrazione italiana.

1. L'atto amministrativo è non solo un atto tecnico-giuridico, ma anche un atto comunicativo con il quale la pubblica amministrazione deve farsi capire e conoscere dai cittadini; atti redatti secondo regole che ne aumentino l'omogeneità e diminuiscano contraddizioni e imprecisioni sono, ovviamente, più comprensibili.
2. La recente pubblicazione del codice dell'amministrazione digitale è l'ultimo atto di un più che decennale percorso normativo che intende sostituire una pubblica amministrazione composta da monadi non comunicanti, con un unico sistema informativo pubblico fortemente interconnesso. In tale quadro l'interoperabilità dei dati è elemento centrale. È di nuovo ovvio che gli standard linguistico-documentari prima ancora che quelli tecnico-informatici faciliteranno tale interoperabilità.
3. La diffusione delle nuove tecnologie in ambito pubblico in presenza di un linguaggio amministrativo farraginoso e oscuro, rischia di rendere ancora meno trasparente l'azione amministrativa, trasferendo la gestione e il controllo dell'informazione, dall'ambito politico-amministrativo a quello tecnico e arrivando persino a mettere in discussione i ruoli d'imparzialità e di garanzia propri della pubblica amministrazione.

Sulla base di queste considerazioni stiamo lavorando ad un progetto che prevede la compilazione di un manuale di regole per la redazione degli atti amministrativi e l'adeguamento di metodi e strumenti per il riconoscimento automatico del linguaggio e per il supporto alla redazione degli atti normativi sviluppati presso l'ITTIG.

2. Manuale per la redazione degli atti amministrativi

Abbiamo realizzato una prima bozza, del tutto provvisoria ed incompleta tentando il coordinamento tra la Direttiva sulla semplificazione del linguaggio dei testi amministrativi del Ministero della funzione pubblica del maggio 2002 e le Regole e Suggestimenti per la redazione dei testi normativi del febbraio 2002¹.

¹ La direttiva è il frutto di un progetto, voluto agli inizi degli anni '90 dall'allora Ministro per la Funzione Pubblica Sabino Cassese che ha prodotto anche un "Manuale di Stile" (Fioritto (a cura di), Manuale di Stile, strumenti per semplificare il linguaggio delle amministrazioni pubbliche, Bologna, Il Mulino, 1997) ed un programma per il controllo della

Per tale coordinamento si è tenuto conto che:

- la direttiva detta alcune raccomandazioni generali che hanno lo scopo di rendere comprensibili e comunicativi gli atti amministrativi, ma si pone solo indirettamente il problema di costruire standard necessari all'interoperabilità dell'informazione;
- diversa invece è la prospettiva del Manuale di drafting normativo che si pone anche la finalità di introdurre standard per rispondere alle necessità d'integrazione e coordinamento degli atti normativi. Tuttavia solo alcune fra le regole di tecnica legislativa sono trasferibili al settore degli atti amministrativi (è ovvio che si possono proporre standard identici per la citazione degli atti normativi, mentre regole peculiari che riguardano ad esempio le modifiche normative non sono applicabili all'ambito amministrativo). Inoltre anche ove tali regole siano applicabili è necessario adeguarle al loro utilizzo per gli atti amministrativi.

Lo sforzo operato di confermare, per quanto possibile, le regole vigenti per gli atti normativi, anche per gli atti amministrativi, risponde peraltro all'esigenza di affermare uno standard il più possibile comune.

Inoltre dall'analisi, appena iniziata, di disposizioni di legge, della prassi e della dottrina abbiamo ricavato regole specifiche per la stesura dell'atto amministrativo che abbiamo inserito nella bozza di manuale.

Abbiamo quindi suddiviso il manuale in tre parti:

1. Il linguaggio degli atti amministrativi.
2. Il rinvio ad altri atti.
3. La struttura degli atti amministrativi.

Tali regole e suggerimenti il cui ambito di applicazione va dagli atti amministrativi su supporto cartaceo a quelli su supporto digitale hanno una triplice finalità: migliorare la comprensibilità e la capacità comunicativa degli atti amministrativi, offrire maggiori garanzie circa la loro legittimità, facilitare il loro trattamento informatico.

2.1 Il linguaggio degli atti amministrativi

Sia la direttiva di semplificazione sia il manuale unificato di tecnica legislativa prevedono regole e suggerimenti per l'uso della lingua italiana nei testi legislativi e amministrativi.

In particolare la direttiva fornisce indicazioni circa l'organizzazione concettuale e strutturale, lo stile e le tecniche di controllo linguistico dei testi.

D'altra parte il manuale unificato di tecnica legislativa propone regole che disciplinano non solo gli aspetti stilistici ma anche quelli morfologici, sintattici e lessicali; ad esempio l'uso dei verbi, delle congiunzioni, l'uso della terminologia tecnico-giuridica, ecc. Tali regole da un lato sono comuni a quelle in vigore per la lingua italiana, dall'altro o ne puntualizzano certi aspetti (vedi ad esempio la regola che disciplina l'uso delle maiuscole²) o se ne distaccano (si veda la disciplina della ripetizione di termini³).

comprensibilità dei testi amministrativi. Il Manuale di tecnica legislativa cui si fa riferimento è il documento "Regole e suggerimenti per la redazione dei testi normativi" voluto dalla Conferenza dei Presidenti dei Consigli Regionali e adottato a partire dal 1991 dalle regioni italiane nella sua versione aggiornata nel 2002. Anche il Parlamento e la Presidenza del Consiglio dei Ministri condividono regole di legistica praticamente identiche emanate nel 1986 e poi aggiornate nel 2001.

² 18. Uso delle lettere maiuscole

Abbiamo dunque incluso queste regole in questa parte del manuale, suddividendole in capi dei quali il primo si occupa dell'efficacia comunicativa del testo amministrativo, dettandone l'organizzazione generale; il secondo descrive le modalità di utilizzo della lingua italiana all'interno dell'atto amministrativo (uso dei segni tipografici, specifiche indicazioni morfologiche, sintattiche e lessicali); il terzo infine suggerisce alcune modalità per la verifica della correttezza del testo.

2.2 Il rinvio ad altri atti

La seconda parte del manuale riguarda soprattutto le tecniche di redazione delle citazioni per il rinvio ad altri atti. Le tecniche di redazione delle citazioni di atti normativi, sono mutate direttamente dal manuale di tecnica legislativa in modo da riproporre anche negli atti amministrativi gli standard già consolidati negli atti normativi.

Quanto alle modalità di scrittura di citazioni di atti amministrativi si è invece proposto uno standard che garantisca l'univocità del rinvio e al tempo stesso la sua intellegibilità.

Lo schema individuato è il seguente:

- a) denominazione formale dell'atto;
- b) autorità emanante;
- c) data;
- d) numero di protocollo;
- e) oggetto.

2.3 La struttura dell'atto amministrativo

Nella parte conclusiva del manuale ci siamo occupati di definire la struttura essenziale dell'atto amministrativo.

La forma e la funzione dell'atto amministrativo richiedono la presenza di alcuni elementi che servono alla sua identificazione e di altri che ne garantiscono l'efficacia giuridica.

Abbiamo identificato e descritto tali elementi e li abbiamo raggruppati in tre parti distinte dell'atto amministrativo in modo da ottenere una organizzazione formale, funzionale e contenutistica idonea ad una formulazione omogenea, chiara e sistematica dell'atto stesso.

La parte iniziale contiene la denominazione dell'atto (ad esempio: decreto, deliberazione, ordinanza, parere, determinazione), l'intestazione che deve indicare l'autorità emanante, il numero di protocollo dell'atto e l'oggetto o titolo, che esprime la funzione e il contenuto dell'atto.

La parte centrale dell'atto contiene il preambolo con l'illustrazione del procedimento che ha portato all'emanazione dell'atto e di quanto ne è emerso, la motivazione che contiene i presupposti di fatto e le

L'uso della maiuscola è prescritto:

- a) all'inizio del testo e dopo ogni punto fermo;
- b) per i nomi propri di persona;
- c) per i nomi propri geografici;
- d) per i nomi di enti od organi individuali.

³ 16. Ripetizione di termini

1. Le singole partizioni dell'atto sono unità autonome del testo.

2. La ripetizione dei termini come pure l'uso di riferimenti normativi completi, sono utili e spesso necessari alla comprensione del testo e all'uso di sistemi informatici. Evitare, pertanto, l'uso di pronomi personali o pronomi e aggettivi dimostrativi se tale uso genera ambiguità. Ripetere invece il termine richiamato quand'è necessario a evitare equivoci.

ragioni giuridiche alla base della decisione ed infine il dispositivo che racchiude la parte propriamente prescrittiva dell'atto amministrativo e contiene, a seconda delle varie tipologie di atto, una manifestazione di volontà, oppure una manifestazione di scienza-conoscenza (es., certificazione, attestazione, verbale di accertamento, ecc.), o di valutazione (es. parere, proposta, valutazione tecnica, ecc.), ecc.

La parte conclusiva dell'atto deve invece necessariamente prevedere il luogo in cui l'atto amministrativo è stato adottato, la data di adozione dello stesso e la sua sottoscrizione.

3. Strumenti di Natural Language Processing

Come detto i vantaggi conseguibili attraverso la marcatura di un testo diventano effettivi solo se si dispone di strumenti efficienti di riconoscimento automatico del testo.

Tali metodi e strumenti sono già stati sperimentati con successo per l'identificazione e la formalizzazione di strutture del testo normativo tanto pare prospettabile un loro utilizzo anche per il riconoscimento di strutture degli atti amministrativi.

A titolo di esempio illustriamo una procedura per il riconoscimento automatico di particolari strutture presenti nel testo normativo; esempio che, tra l'altro, si riferisce alla citazione testuale esplicita di atti normativi che è una struttura ricorrente anche negli atti amministrativi.

In sintesi, la metodologia d'analisi adottata è la seguente:

- creazione di uno o più modelli, sulla base di strutture prescritte da regole o definite dalla teoria e dalla prassi, per descrivere il microtesto da analizzare, e il suo rapporto all'interno dell'intera struttura testuale;
- riscrittura di tali modelli in regole di produzione;
- scelta del campione;
- analisi sul campione, esame dei risultati e definizione di nuovi modelli descrittivi, a partire dai risultati individuati;
- ulteriore verifica dei modelli così ottenuti su un corpus più ampio e trasferimento dei modelli in un sistema di riconoscimento automatico del linguaggio.

Lo strumento di cui ci si è avvalsi, particolarmente flessibile e configurabile è il sistema di estrazione dell'informazione Sophia 2.1, che consente la formalizzazione di regole e modelli specifici.

In particolare con il software suddetto si è proceduto all'analisi e marcatura di un primo campione di testi normativi secondo le fasi seguenti:

- normalizzazione del testo in ingresso, marcando adeguatamente tutte quelle strutture e quei segmenti testuali riconoscibili sulla base dei caratteri, ovvero senza ricorso a consultazione del lessico-dizionario;
- analisi lessicale (categoria sintattica) e morfologica (tratti di flessione) del testo in ingresso;
- disambiguazione della categoria sintattica delle parole (*Part of Speech Tagging*);

- analisi sintattica parziale (denominata *chunking*), tesa ad identificare i gruppi sintattici minimi presenti nel testo in ingresso e a raggrupparli in costituenti;
- analisi semantica e identificazione delle strutture concettuali rilevanti nel testo d'ingresso;
- conversione del documento analizzato dal formato Microsoft® Word (HTML, RTF, txt, ecc.) in formato XML, secondo la DTD stabilita.

3.1. La citazione normativa

Abbiamo individuato quali elementi costitutivi e distintivi della citazione normativa:

1. la **Partizione** che indica qualsiasi parte esplicitamente contrassegnata da una particolare espressione grafica in cui si suddivide l'atto (ad es. articolo, comma, lettera, capo, titolo, ecc.) con il relativo segno grafico di enumerazione (es. 3, c), ottavo, A, ecc.);
2. l'**Atto (nome+data+numero)**. Per Atto s'intende l'identificatore univoco dell'atto normativo richiamato, che, di regola, si esprime con il nome ufficiale dell'atto (es. legge, decreto-legge, decreto ministeriale, ecc.), la data di promulgazione o emanazione, il numero d'ordine assegnato all'atto.

Innanzitutto la citazione deve essere univoca per individuare quello, e solo quello, specifico atto a cui si riferisce. Quasi sempre questa univocità si ottiene riportando tre informazioni:

il *nomen* dell'atto (che indica la categoria a cui l'atto appartiene);

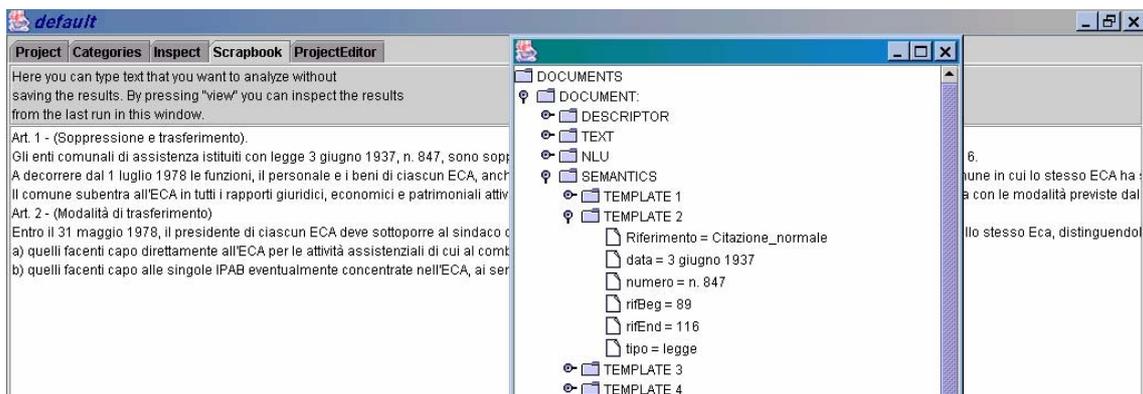
la *data*, di solito, quella di promulgazione o di emanazione (il formato standard è: gg (in cifre)/mese (in lettere)/aaaa (in cifre));

il *numero* che identifica l'atto normativo (ha un massimo di quattro cifre).

A titolo d'esempio vediamo come si esprime un modello di citazione nel modulo di formalizzazione delle struttura semantica, nel quale si utilizzano elementi definiti dalle regole di produzione dei livelli inferiori (lessicale, morfologico e sintattico).

Citazione_normale	[M-NOME-ATTO]+ DATE:vdata+ PUNCTX+ NUM-ATTO:vnum	{actionNEW(vltipo:Riferimento:"Citazione_normale", vlqualifica:qualifica_tipo_atto:alllist, vltipo:tipo:alllist, vlautorita:autorita:alllist, vdata:data:all, vnum:numero:all)}
-------------------	---	--

Tale formalizzazione consente di identificare le citazioni e di generare in output il template riprodotto qui sotto.



Gli elementi che compongono la regola sopra indicata necessitano di alcuni approfondimenti.

In primo luogo bisogna notare che la regola suddetta è stata formalizzata prevedendo come obbligatori tutti gli elementi della citazione: il nome dell'atto (M-NOME-ATTO), la data (DATE), la virgola (PUNCTX) tra la data e il numero dell'atto e lo stesso numero dell'atto (NUM-ATTO). Abbiamo cioè scritto la regola "Citazione_normale" in modo da avere corrispondenza piena con le regole di tecnica legislativa.

3.2 Conversione in XML

Una volta identificati nel testo gli elementi che costituiscono le citazioni la marcatura in XML di questi elementi è pressoché immediata.

Il sistema di analisi del documento memorizza infatti le posizioni degli elementi riconosciuti all'interno del testo.

Si può quindi generare facilmente un nuovo documento che contiene il testo marcato, come è esemplificato nel frammento riportato sotto:

```
<rif articolo="1" autorita="regionale" comma="2" data="14 ottobre 1999" nome="legge" numero="52" rule="Partiz_invertita"> I comma 2 dell'articolo 1 della legge regionale 14 ottobre 1999, n. 52 </rif>
```

4. Conclusioni

La comprensione del testo amministrativo da parte della macchina, attraverso l'introduzione di standard e l'uso di metodologie di NLP introduce a varie applicazioni per il trattamento e la comunicazione degli atti stessi, quali:

1. introdurre modelli di redazione e valutazione dei provvedimenti comunali;
2. fornire avanzati strumenti informatici di supporto alla produzione degli atti;
3. agevolare ai cittadini la consultazione degli atti del Comune di Firenze.

Le applicazioni sopra accennate sono in via di progettazione nell'ambito del progetto "REgole chiaRE", che l'Istituto di Teoria e Tecniche dell'Informazione Giuridica del CNR ha predisposto in collaborazione col Comune di Firenze.