

Capitolo A05 procedure di supporto alla esposizione

Contenuti delle sezioni

- a. motivazioni e compiti p. 4
- b. prestazioni di controllo e operazioni diagnostiche p. 6
- c. generazione dei componenti derivabili di una esposizione p. 8
- d. manovre per sostenere il controllo del testo p. 10
- e. scelte per i programmi e per le strutture di dati p. 12
- f. strutture di dati per il testo-M p. 14
- g. manovre per costruire il testo-M p. 17
- h. prospettive di sviluppi ulteriori p. 19

19 pagine

A050.01 In questo capitolo si considera l'*esposizione* come un testo in fieri sul quale si sperimentano modalità per la presentazione di nozioni della matematica e delle sue applicazioni che favorisce la possibilità di servirsi di automatismi in grado di agevolare lo sviluppo della presentazione stessa.

Nelle pagine che seguono si espongono le considerazioni che hanno condotto a definire le suddette modalità e le procedure automatiche, che chiamiamo **procedure di supporto**, o anche procedure proeXp” che vengono messe a punto per sostenere lo sviluppo dell'*esposizione*.

Queste procedure, più in generale, vorrebbero supportare l'evoluzione di testo con contenuti dell'area STEM (science, technology, engineering, mathematics), ossia testi nei quali si richiedono definizioni e argomentazioni accurate e complete, nei quali si incontrano formule matematiche e algoritmi che devono poter essere interpretati senza ambiguità.

Per questi testi in corso di sviluppo si chiede anche la possibilità di agevolare il riutilizzo di loro parti per più tipi di varianti, ad esempio per presentazioni di nozioni specialistiche che risulti opportuno tenere collegate a nozioni generali oppure per presentazioni semplificate per scopi didattici e divulgativi, ma con la possibilità di collegarsi a ben definiti discorsi di approfondimento.

A050.02 Per sommi capi, le intenzioni delle procedure automatiche accennate, intendono favorire la coerenza lessicale e strutturale del testo corrente, la rilevazione di certi tipi di mancanze del testo, e la effettuazione di alcuni tipi di modifiche, di unificazioni e di differenziazioni di elementi espositivi. Si tratta di operazioni che possono richiedere a un redattore di localizzare elementi del testo da ricercare entro qualche migliaio di pagine.

Un'altra finalità delle procedure di supporto consiste nel generare automaticamente parti del testo logicamente ricavabili dalle rimanenti come: la tavola dei contenuti, gli indici delle occorrenze di termini e di simboli specifici, le citazioni delle personalità richiamate.

Le procedure di sostegno annunciate concretamente consistono in programmi scritti in un sottoinsieme ben delimitato del linguaggio C, ma con aggiunta di qualche semplice prestazione del linguaggio C++, che richiedono competenze di programmazione molto limitate, anche solo quelle riguardanti il linguaggio c++ presentate nei capitoli da B80 a B88.

Questi programmi vengono messi a punto con il sistema di sviluppo Microsoft Visual Studio e possono essere utilizzati su un qualsiasi personal computer. Si tratta in sostanza di strumenti che richiedono risorse molto ridotte.

A050.03 Le pagine che seguono vogliono anche suggerire come si possono costruire procedure che consentano di sostenere lo sviluppo di presentazioni di argomenti scientifico-tecnologici diverse da MATeXp, ma con caratteristiche simili.

Più precisamente si pensi a presentazioni redatte nel linguaggio T_EX contenenti formule matematiche e diagrammi schematici riguardanti andamenti quantitativi e comportamenti strutturati, che si servono di termini accurati e precisamente definiti, che hanno una ben definita organizzazione con ripartizioni di diversi livelli, come quella riguardante tomi, capitoli, sezioni e paragrafi.

Qualcuna delle considerazioni che seguono potrebbe servire anche a testi redatti in un altro linguaggio, purché organizzato con una versione sorgente esprimibile con caratteri ASCII o UNICODE e una versione leggibile normalmente derivabile dalla precedente. Vediamo queste caratteristiche nei linguaggi HTML e XML e il linguaggio dei sistemi Wikimedia.

Per la suddetta intenzione di prospettare possibilità diverse da MATeXp in questo capitolo ci si indirizza spesso a una presentazione che potrebbe essere una variante di MATeXp e che per genericità denotiamo con la scrittura “esposizione”.

Nel seguito compaiono spesso scritte di frammenti di programmi; ad esse e solo per esse ci serviamo di questa `font teletype` .

A05 a. motivazioni e compiti

A05a.01 Il testo di MATeXp o di altre esposizioni di argomenti dell'area STEM che intendono essere giudicati attendibili e utili per applicazioni concrete, si può considerare una struttura informativa che presenta parti tendenzialmente discorsive e parti con scopi direttamente o indirettamente operativi che si servono di numerose convenzioni specialistiche.

Una tale struttura informativa, per essere utilizzabile deve osservare un complesso di regole piuttosto ampio e articolato.

Quando un testo di questo genere è in corso di sviluppo e quando si vuole che i suoi contenuti siano adattabile a possibili nuove esigenze, deve seguire regole da considerare non immutabili, ma che possano essere in qualche misura modificate per adattarsi a prospettive poco diverse dalle originarie.

Potrebbe essere richiesto di adattare i contenuti a esigenze definite con maggiore precisione o all'opposto a esigenze più generali; può accadere che si richieda una variante più mirata a sostenere tipi di applicazioni emersi recentemente e diverse da quelle prevalenti nel passato; potrebbe essere necessario rendere una esposizione più facilmente fruibile, oppure meno dipendente da premesse veosimilmente ignorate da nuovi destinatari; all'opposto potrebbe essere richiesta una versione da utilizzare per applicazioni urgenti che dà per scontate premesse che erano state attentamente analizzate.

Qualche nuova esigenza esterna potrebbe richiedere adattamenti sistematici di termini e di notazioni; potrebbero rendersi opportuni cambiamenti in grado di consentire una più agevole manutenzione.

Potrebbero anche essere richieste modifiche che consentono di ottenere risultati che si giudicano ragionevolmente raggiungibili dopo che sono state rese superabili difficoltà precedentemente giudicate fuori portata.

Il preoccuparsi di tutte queste varianti si può giustificare con la previsione che nel futuro il crescere dell'importanza delle innovazioni tecnologiche comporteranno una forte crescita della domanda di istruzione su contenuti dell'area STEM.

A05a.02 Per ottenere un buon rispetto delle regole risulta utile per il loro mantenimento nel tempo e risulta opportuno per lo sviluppo complessivo di una esposizione ricorrere a procedure computazionali in grado di analizzare i testi per segnalare i mancati rispetti delle regole che possono essere classificati come refusi, mancanze, incoerenze, ambiguità e inutili ripetizioni.

Va anche segnalato che i supporti computazionali a un corpo informativo disponibile in forma digitale e che intende rispettare regole chiare, come cerca di essere MATeXp, consente di facilitare in misura rilevante la sua riutilizzazione per varianti degli obiettivi, in particolare per scopi didattici e divulgativi.

Questo ultimo genere di prestazioni riguarda MATeXp solo per le sue possibili riprese per finalità ben delimitate, ma conviene prospettarle, in quanto la loro implementazione può avvalersi di parecchi moduli di programma e di parecchi schemi operativi che devono far parte della strumentazione computazionale riguardante le finalità sopra accennate.

A05a.03 Attualmente si procede nello sviluppo delle procedure prospettate servendosi di alcuni source editors liberamente disponibili, PSpad e Notepad++, e con i preannunciati programmi che si stanno mettendo a punto specificamente per MATeXp nel linguaggio C++ utilizzato soprattutto con modalità procedurali al livello della applicazione desktop e avvalendosi solo di librerie di uso comune.

Vediamo in breve le regole che vengono fatte rispettare e gli arricchimenti da generare automaticamente.

Il testo viene organizzato in ripartizioni diversamente comprensive, capitoli, sezioni e paragrafi che si presentano seguendo un ordine semirigido, ossia corrispondente a sigle caratterizzanti che rispettano un semplice ordine alfabetico o numerico, ma che si permettono lacune con il fine di non ostacolare successive modifiche integrative.

In ciascuno dei paragrafi si distinguono chiaramente parti discorsive scritte in un italiano (o in altra lingua naturale) contenente tecnicismi piuttosto specifici,

parti riguardanti formule racchiuse tra \$ e \$ oppure tra \$\$ e \$\$,

nidi chiaramente delimitati come `{\dn termine}`, `\YA{autore}`, `\Yw{opera}`, `\IL{rinvio interno}`... .

Le parti discorsive, oltre alle regole della lingua naturale, è necessario che rispettino attentamente la precisione lessicale.

Si devono inoltre rispettare le regole della consequenzialità della presentazione.

In particolare ogni nozione specifica che non si supponga ampiamente riconosciuta dovrebbe essere usato solo dopo essere stata ben definita; a questa norma spesso si derogare per mantenere la scorrevolezza della presentazione, ma in tali casi si dovrebbe esplicitare la deroga stessa e si dovrebbero segnalare i punti del testo o i riferimenti esterni nei quali si possono trovare indicazioni in grado di colmare la lacuna di comprensione che si è rischiesta.

A05a.04 In un testo sorgente per T_EX tutto quello che in una linea segue un carattere % che non sia preceduto da \ non influisce sulla quanto viene reso leggibile.

Queste parti del testo sorgente invisibili per il lettore possono essere utilizzate dal redattore per segnalazioni che lo possono aiutare in vari modi. Scopi analoghi si possono ottenere con particolari macro T_EX anch'esse invisibili inseribili in ogni punto che sarà reso leggibile. In particolare sono disponibili macro definite con richieste della forma

$$\backslash\text{def}\backslash\text{xyy}\{\}$$

Si possono marcare con stringhe invisibili che conviene scegliere con attenzione punti del testo da sviluppare, da accertare, da precisare, da ricollocare o da perfezionare in momenti successivi.

Risulta anche opportuno marcare adeguatamente elementi del testo che si prevede debbano essere aggiornati con regolarità in relazione con il procedere dell'ampliamento dei contenuti.

Oltre a queste segnalazioni utili all'evolversi della redazione, nelle parti invisibili si possono inserire stringhe che vengono utilizzate da procedure che scorrono il testo sorgente per ricavare indicazioni di vari generi.

Con queste finalità si possono inserire in ogni posizione del testo sorgente macro con nomi per i quali si chiede soltanto che non interferiscano con le altre stringhe efficaci che consentono di individuare elementi trattabili da procedure prXp per scopi simili a quelli sopra accennati.

A05 b. prestazioni di controllo e operazioni diagnostiche

A05b.01 Le prestazioni di controllo delle procedure di supporto si basano sulla individuazione nel testo sorgente di elementi qualificanti e sulla verifica della loro accettabilità lessicale e formale.

Le verifiche sui diversi tipi di elementi che si trovano nel testo si effettuano in modi che dipendono dalle caratteristiche degli elementi.

Alcune verifiche si effettuano con algoritmi tendenzialmente semplici riguardanti stringhe di lunghezza contenuta; questo accade ad esempio per le coordinate dei paragrafi del testo aventi le forme “Tnn”, “TnnS” e “TnnSmm”, riconoscibili facilmente.

Per altri tipi di elementi la verifica della accettabilità si realizza stabilendo se la stringa elemento compare o meno in un lungo elenco di parole del linguaggio della matematica o di stringhe convenzionali, in particolare le macro del linguaggio Plain T_EX e le macro definite specificamente per MATeXp.

Questi elenchi di stringhe di riferimento vanno definiti con accuratezza e devono essere aggiornati in accordo con le modifiche che si vanno effettuando sui testi, soprattutto con gli arricchimenti complessivi.

Si può pensare che per sviluppare esposizioni controllabili con procedure del genere proExp convenga partire da un insieme collaudato di elementi e di tipi di elementi che si possa ritoccare e far crescere con una certa facilità mentre si procede con le modifiche e con l’ampliamento del testo, avendo cura che risulti sempre relativamente agevole far entrare nel testo elementi di tipi nuovi lessicalmente non previsti o formalmente trattabili con nuovi algoritmi che devono tener conto solo di su stringhe concise e ben distinguibili.

In questi ampliamenti, evidentemente, si deve curare la compatibilità con il lessico e con gli algoritmi precedenti.

A05b.02 Presentiamo le manovre riconducibili al controllo della correttezza che sono state prese in considerazione, cominciando con quelle che chiamiamo “estrazioni”, manovre che consistono nello scorrimento e nell’analisi dei files nei quali è registrato il testo sorgente alla ricerca degli elementi da sottoporre a verifica.

Estrazione dei titoli di capitoli e sezioni al fine di verificare il loro ordinamento e di tenere sotto controllo il loro complesso.

Estrazione dei nidi della forma $\{\backslash\text{dn}\textit{termine}\}$, ciascuno accompagnato dalla relativa coordinata di paragrafo della forma “TnnSmm”.

Una tale operazione la chiamiamo anche “estrazione e localizzazione” del nido e serve a tenere sotto controllo il complesso dei termini più significativi e ad evitare che tra di essi si abbiano conflitti di omonimia, ambiguità di sinonimia e varianti che possano generare confusione quando vengono adottati senza gli opportuni confronti con scelte potenzialmente conflittuali.

Il termine “estrazione e localizzazione” viene usato anche per denotare altre operazioni simili che stiamo per incontrare.

Estrazione delle occorrenze di notazioni precedentemente marcate nelle parti invisibili, anch’esse accompagnate dalle relative coordinate; servono a controlli simili a quelle sui termini giudicati significativi.

Estrazione e localizzazione delle citazioni di autori o di altre personalità in quanto elementi del testo della forma $\backslash\text{Ya}\{\textit{nome}\}$; di questi elementi si intende controllare l’ortografia e la reperibilità di informazioni che di essi si trovano sul Web; questi tipi di dati vengono esaminati nel fascicolo X15.

Estrazione e localizzazione delle citazioni di opere fornite da elementi della forma $\backslash\text{Yw}\{titolo\}$; questi dati servono solo a controllare la loro presenza nel capitolo X16.

Estrazione dei rinvii della forma $\backslash\text{IL}\{coordinata\}$ e registrazione della relativa coppia $\langle \text{paragrafo rinviante} , \text{paragrafo citato} \rangle$.

Per ciascuna di queste coppie si verifica la correttezza del collegamento, ossia l'effettiva esistenza del paragrafo bersaglio.

Operazioni e scopi simili per i rinvii alle sezioni e ai capitoli.

Individuazione delle stringhe alfanumeriche esterne ai nidi ... e

...

e simili, stringhe da considerare lessicalmente accettabili nel resto della attuale esposizione; deve quindi seguire la verifica della loro presenza nell'apposito elenco generale delle stringhe accettabili mantenuto al fine di scovare i possibili refusi nel testo e di avere una visione, ulteriormente analizzabile e precisabile del lessico adottato in relazione al complesso di conoscenze trattato dalla esposizione in esame.

Individuazione delle macro della forma $\backslash\text{stringa alfabetica}$ e per ciascuna di esse verifica se sia presente o meno nell'elenco di quelle accettabili.

Vengono esaminate distintamente le macro del sistema T_EX e le macro definite dal redattore. Si deve garantire che non si abbiano conflitti di omonimia, e che non vi sia sinonimia tra le macro del redattore; inoltre è opportuno che queste ultime vengano usate in modo coerente, evitando le confusioni, possibili soprattutto quando si aggiungono ai contenuti nuovi argomenti con complesse relazioni con le macro precedenti.

A05b.03 Passiamo ora in rapida rassegna le operazioni di diagnosi dei files sui quali è registrato il testo finalizzate a verificare e mantenere il rispetto della loro organizzazione e la loro coerenza, almeno al livello formale.

Scorrimento dei files forniti da un dato elenco.

Scorrimento di una data porzione di un dato file.

Scorrimento di una riga di un file fino alla fine oppure fino alla prima occorrenza del segno “%”.

Controllo in un capitolo della coerenza fra indice iniziale e sequenza delle sezioni, compresa verifica delle coincidenze dei titoli.

Controllo dell'ordinamento delle sezioni e dei paragrafi all'interno di un capitolo con la segnalazione delle eventuali scorrettezze.

Estrazione da alcuni capitoli dei titoli dei capitoli e delle sezioni per assicurare l'aggiornamento dei contenuti dei capitoli A00 e X10.

Individuazione in alcuni capitoli dei rinvii T_mnS_mm, verifica della presenza del bersaglio – paragrafo, sezione o capitolo – e raccolta delle coppie $\langle \text{paragrafo rinviante} , \text{ripartizione bersaglio} \rangle$.

In prospettiva esami complessivi della consistenza dei vari argomenti trattati e della adeguatezza dei collegamenti tra i vari argomenti basati su occorrenze di termini, correlazioni fra termini e collegamenti fra le parti.

A05 c. generazione dei componenti derivabili di una esposizione

A05c.01 Elenchiamo rapidamente le operazioni che servono alla generazione di parti dell'esposizione desumibili da altre parti.

Per ogni capitolo costruzione dei rinvii dalle linee del capitolo ai corrispondenti paragrafi.

Estrazione di termini dalle righe del testo che iniziano con

`%;T TnnSmm`

e controllo dell'esagramma coordinata.

Estrazione riga notazione controllo riga notazione e correttezza della sua coordinata

Estrazione di notazioni da file delle righe che iniziano con

`%;Na TnnSmm`

e controllo dell'esagramma coordinata.

Estrazioni di notazioni analoghe da righe che contengono le scritte

`%;Nc , %;Nd , %;Ne , %;Ni , %;Nm`

Costruzione di righe per l'indice dei termini

Costruzione di righe per l'indice delle notazioni

Modifica di intere porzioni del testo ottenuta con richieste di eliminazione o di inserimento di intere sezioni di interi paragrafi.

A05c.02 Per concludere la presente sezione accenniamo ad altre due prospettive per le attività di supporto sviluppo di MATeXp o di esposizioni analoghe.

La prima riguarda la traduzione del suo testo in altra lingua.

La traduzione di materiali matematici e computazionali di una certa ampiezza può essere significativa, soprattutto in questo periodo di rapida crescita degli interessi nei confronti delle svariate innovazioni scientifico-tecnologiche e delle loro conseguenze socio-economiche.

Inoltre le stesse recenti innovazioni nel settore della traduzione automatica hanno resi disponibili strumenti di sostegno per le traduzioni auspiccate che le rendono fattibili più agevolmente, più rapidamente e con maggiore consapevolezza di quanto molti potessero fare nel passato. nel passato.

Si osserva inoltre che molti componenti di una esposizione possono essere inseriti in una sua traduzione senza cambiamenti (formule, algoritmi, schemi, ...) o in seguito a trasformazioni che possono essere effettuate con il supporto di automatismi da sviluppare appositamente.

Questi sviluppi di procedimenti e di dati dovrebbero essere messi a punto nell'ambito di una visione complessiva delle componenti (essenzialmente termini ed espressioni) delle formulazioni delle nozioni matematiche e delle operazioni di gestione e trasformazione dei testi.

Qualche tentativo è stato avviato, sempre a un modesto livello artigianale, per trasformare MATeXP nella lingua inglese.

Si è osservato che gli attuali traduttori automatici per essere utilizzati efficacemente richiedono operazioni preparatorie e di finitura assai attente, ma anch'esse facilitate da operazioni di elaborazioni di testi che possono essere studiate sistematicamente.

Inoltre emerge l'opportunità di disporre di dizionari bilingue che si attengono a scelte lessicali di una certa portata che vanno decise per ciascuna delle lingue in causa, e che sarebbe auspicabile possano avvalersi di contributi di gruppi di esperti competenti.

A05c.03 Una seconda prospettiva riguarda l'adattamento del testo ad altri linguaggi per la comunicazione che prevedano la agevole disponibilità di un linguaggio sorgente trattabile con programmi procedurali diffusi, come C++ e tanti altri (Java, Python, PHP, ...).

Una prima alternativa, riguardante un cambiamento relativamente piccolo, consiste nell'uso di LaTeX invece di Plain T_EX.

Una seconda alternativa per il rimpiazzamento di Plain T_EX è un linguaggio per il Web come HTML. Questo linguaggio consente una organizzazione della ipertestualità dei contenuti molto più efficace; in questa direzione si aprono nuove possibilità, accompagnate dai relativi problemi.

Per passare ad HTML si pongono problemi riguardanti principalmente le formule matematiche, sia come porzioni circoscritte di frasi discorsive che come brani autonomi di una certa rilevanza.

In questa direzione sembra siano richiesti interventi, anche impegnativi, del redattore e di una persona che dovrebbe conoscere sia i linguaggi di comunicazione coinvolti che i contenuti da tradurre.

Conviene osservare che le operazioni di traduzione di una esposizione potrebbero essere accompagnate da modifiche di parti dello stesso testo originale che possono costituire miglioramenti nelle formulazioni indotte dal confronto fra le loro versioni in lingue diverse.

Questi miglioramenti di ritorno possono essere simili a quelli che possono provenire da ogni ampliamento di una esposizione in relazione alle critiche dei contenuti che possono derivare da ogni ampliamento della visione delle nozioni che presentano molteplici collegamenti con altre nozioni rilevanti, ovvero che possono essere esaminate fruttosamente da vari punti di vista.

A05c.04 Altri strumenti di comunicazione meritano di essere presi in considerazione.

Uno molto interessante è costituito dai sistemi Wiki: anch'essi consentono di accostare le nozioni con prestazioni ipertestuali versatili che presentano indubbi vantaggi, ma per contro richiedono di precisare informazioni e collegamenti necessari ai meccanismi non richiesti da procedimenti di comunicazione che trattano materiali di studio leggibili che non hanno la possibilità di adattarsi alle esigenze del fruitore nel corso della sua lettura.

In una direzione ben diversa si possono cercare versioni di MATeXp in sistemi di comunicazione di larghissima diffusione, ad esempio in un word processor come Word di MS Office o simili; maggiori indicazioni su questi strumenti non siamo in grado di fornire.

Tutte le varie alternative indicate sembrano avere caratteristiche interessanti per ottenere da un testo esteso e disciplinato estratti circoscritti ragionevolmente controllabili da una sola persona e in tempi relativamente ridotti.

A05 d. manovre per sostenere la controllabilità del testo

A05d.01 Attualmente per rendere il testo di MATeXp controllabile con una buona sicurezza e senza eccessivo impegno umano si agisce soprattutto sul testo sorgente in due modi.

(1) Si mantiene il testo rispettoso di precise regole riguardanti la struttura delle sue ripartizioni, l'impiego di un lessico controllato e l'adeguatezza delle notazioni che si adottano.

(2) A quanto è destinato ad essere reso visualizzabile, attraverso files con estensione `.pdf` come sequenza di pagine leggibili su uno schermo o su carta, si aggiungono commenti nelle parti finali delle linee, quelle che seguono la prima occorrenza del carattere `%`; questi commenti è opportuno siano anch'essi regolati da criteri di omogeneità e chiarezza formale, in quanto devono essere tenute presenti con chiarezza dai redattori nei tempi lunghi della redazione di un testo molto meno scorrevole di quelli che si incontrano nella gran parte delle narrazioni e in buona parte dei testi con fini culturali.

Questi commenti, che vengono trascurati dal rendering nei files `.dvi` e `.pdf` ottenibili dal programma traduttore WinShell adottato, hanno lo scopo di facilitare gli interventi automatici sul testo sorgente eseguibili mediante comuni editori di testi (PsPAD, Notepad++) e mediante altri programmi di ampia disponibilità (Excel e Regex), in grado di effettuare i più generici controlli di omogeneità e coerenza dei contenuti, controlli che è opportuno rivolgere anche agli stessi commenti del redattore.

Essi inoltre servono a stabilire quali occorrenze di termini, di notazioni, e di altre stringhe con ruoli specifici devono comparire negli indici dei riferimenti interni e consentono una loro elaborazione relativamente agevole per la generazione automatica dei fascicoli destinati a contenere gli indici stessi.

A05d.02 Esaminiamo brevemente i generi delle manovre previste.

Analisi della struttura dei files-capitoli e relativa diagnosi.

Analisi statistica dei files e presentazione dei conseguenti risultati.

Trasformazioni di files `.tex` in files `.tex` per migliorarne alcune caratteristiche formali.

Produzione files di supporto `.txt` e `.htm`

Generazione di indici da sottoporre a postproduzione o definitivi, ossia direttamente pubblicabili sul Web.

A05d.03 Sono previste manovre dei tipi seguenti.

Controllo della organizzazione complessiva dei files-capitoli; prevede la verifica della non ambiguità e dell'ordinamento delle sigle che caratterizzano le ripartizioni in capitoli, sezioni e paragrafi.

Controllo delle parole, grosso modo le stringhe alfanumeriche, che si trovano nei testi, con la segnalazione di quelle escluse dal previsto elenco di riferimento; per ciascuna di esse sarà deciso dal redattore se inserirla nell'elenco di quelle da considerare lecite, oppure la correzione della occorrenza trovata.

Controllo delle macro trovate e verifica della appartenenza di ciascuna loro occorrenza all'elenco delle macro \TeX , oppure all'elenco delle macro del redattore; nel caso di assenza si ha la correzione dell'occorrenza, oppure il suo inserimento nell'elenco che la riguarda.

Controllo dei termini contenuti nei nidi `\tGatermine` consistente nella verifica della loro appartenenza al relativo elenco, che nel caso di termine nuovo viene seguita dalla sua accettazione doppiamente dalla correzione della occorrenza.

Controllo delle notazioni evidenziate invisibilmente come significative; per ciascuna notazione viene precisato il tipo e verificata la sua presenza nell'elenco di quelle previste e nel caso di assenza viene decisa la correzione o l'inclusione nell'elenco.

Controllo delle occorrenze delle stringhe dei tipi previsti nell'elenco AbbrSpecVar nel file X13 seguito, nel caso di assenza, dalla correzione o dell'inclusione.

Controllo dei nomi di persone che compaiono nei nidi $\backslash\text{Ya}\{nome\}$ seguito, nel caso di assenza, dalla correzione o dell'inclusione.

Controllo delle citazioni delle opere che compaiono nei nidi $\backslash\text{Yw}\{citazione\}$ seguito, nel caso di assenza, dalla correzione o dell'inclusione.

A05d.04 Presentiamo le operazioni che toccano le strutture informative di riferimento, relativamente semplici, che vengono richiamate da uno o più tipi di manovre.

Gestione di un ordinamento di tipo lessicografico mediante un automa a stati finiti.

Scorrimento di una lista di files-capitoli e delle corrispondenti linee.

Accettazione, qualificazione, eventuale segnalazione diagnostica di una linea di un testo sorgente ed eventuale sua correzione.

Controllo della nonambiguità e del rispetto dell'ordinamento di una sezione o di un paragrafo.

Ricerca della presenza di una stringa in un elenco che segue un preciso ordinamento lessicografico, seguita da segnalazione di assenza.

Implementazione di un accettatore-trasduttore a stati finiti in grado di sottoporre a diagnosi e a trasduzioni stringhe estratte dal testo sorgente.

Gestione di elenchi che seguono ordinamenti lessicografici definiti da automi a stati finiti (lessemi e tokens) con numeri di occorrenze.

Gestione di un elenco KWIC.

A05 e. scelte per i programmi e per le strutture di dati

A05e.01 Per denotare i programmi che si stanno mettendo a punto per il controllo del testo di MATeXp useremo anche la sigla MTM, acronimo di Mathrmatical Text Manipulation, oppure la sigla MMMC, per man machine math control; in essa man machine math va intesa come “matematica trattabile sia da persone umane che da macchine”.

Questi programmi, come si è detto, sono scritti in un sottoinsieme ridotto del linguaggio C++ come “desktop application”, ossia come software applicativo che come piattaforma per la sua realizzazione richiede un semplice computer di tipo desktop.

Si tratta di programmi poco interattivi (anche se una maggiore interattività sarebbe auspicabile), che prevedono un file di governo che precisa le finalità di una esecuzione, espressi seguendo regole sintattiche limitatamente elastiche, i files da elaborare e i files da generare; si possono avere files da aggiornare che compaiono tra i files da leggere e tra i files di uscita.

Tra i files da leggere spesso si trova un file .txt contenente un elenco dei files-capitoli con estensione .tex che si devono esaminare o modificare.

Tra i files da generare si trova normalmente un file di log di estensione .txt nel quale, per ciascuna delle esecuzioni del programma in esame, vengono registrati i contenuti del file di governo e gli esiti di rilievo della esecuzione effettuata.

Per molte manovre tra i files da generare si trova un file .txt destinato a contenere le segnalazioni delle scritture illegali, delle configurazioni erranee o ambigue e delle configurazioni da chiarire che si sono trovate nei files-capitoli che il governo aveva richiesto di esaminare.

A05e.02 Per avere programmi MTM che possono operare con una buona efficienza anche sui computers più comuni, si prevede di tenere una versione del testo in memoria corredato da un robusto sistema di puntatori che lo rendono una struttura di dati gestibile con efficienza, sia in termini di velocità esecutiva, che nella prospettiva di rendere disponibili procedure di analisi e di correzione facili da progettare e gestire.

Conviene quindi distinguere tra un testo-L, testo sorgente in files .tex leggibili da un addetto e un testo-M equivalente al sorgente consistente nella struttura di dati sopra accennata.

Per ottenere la suddetta efficienza di gestione sono previsti anche data sets ausiliari organizzati in modo da poter essere utilizzati in sintonia con il testo-M.

Va anche segnalato che questi data sets, che chiamiamo “data sets di riferimento”, richiedono di essere accuratamente tenuti aggiornati.

A05e.03 Buona parte dei data sets di riferimento sono previsti lentamente variabili, in quanto tengono conto di scelte organizzative complessive delle nozioni che vengono trattate, scelte che vengono modificate sono in seguito a decisioni di livello strategico per lo sviluppo di MATeXp.

Presentiamo a grandi linee questi data sets.

Elenco dei files .tex contenenti il testo sorgente; in esso si distinguono i files ritenuti stabili da quelli che potrebbero essere notevolmente modificati.

Elenco delle macro TeX.

Elenco delle macro definite dal redattore appositamente per MATeXp.

Elenco dei tipi di files ausiliari previsti.

Struttura informativa riguardante le varie manovre eseguibili con le relative specifiche che per ogni esecuzione devono essere precisate ed essere inserite mediante loro opportune codifiche nei files di governo; le specifiche delle varie manovre devono comprendere le specificazioni dei files di ingresso e di uscita che ogni manovra prevede.

La struttura informativa delle manovre serve per la interpretazione delle richieste contenute nei files di governo, richieste che devono essere esaminate e giudicate legali prima di diventare esecutive.

Repertorio dei messaggi diagnostici previsti (attualmente in un italiano non esente da anglicismi e traducibili senza difficoltà.

A05e.04 Presentiamo, ancora a grandi linee i data sets di riferimento che devono essere aggiornati in seguito di ogni esecuzione che comporta modifiche del testo sorgente di MATeXp.

Elenco di lessemi del testo sorgente, ossia di parole alfanumeriche e di tokens contenenti frammenti-\$\$.

Elenco delle macro, distinguendo le macro \TeX dalle macro del redattore.

Elenco dei termini che compaiono nei nidi-dn.

Elenco delle notazioni dei diversi tipi segnalate dai marcatori %;N.

Elenco delle stringhe chiamate AbbrSpecVar costituente il file X13.

Elenco delle persone citate nei nidi-Ya.

Elenco dei lavori citati nei nidi-Yw.

Codifiche dei classificatori delle linee del testo sorgente

Codifiche degli accettatori-trasduttori riguardanti le righe del testo sorgente e i vari tipi dei frammenti contenuti in queste righe.

Codifiche dei modificatori elementari dei lessemi.

Elenco dei paragrafi contenuti in una lista di files-capitoli con le relative posizioni iniziale e finale.

A05 f. strutture di dati per testo-M

A05f.01 In questa sezione presentiamo le strutture informative, le operazioni e i procedimenti che compaiono nei programmi MTM procedendo in modo tendenzialmente schematico e uniforme.

Ci serviamo di vari termini e codici definite ad hoc, cercando di essere precisi ed esaurienti. Inoltre cerchiamo di essere vicini alle scelte del linguaggio C++.

Per ogni array che incontreremo in seguito, se il suo identificatore è `arr`, denoteremo naturalmente con `arr[i]` il suo componente nella posizione `i`, con `arr[..Larr]` la sequenza dei suoi componenti dalla posizione 0 iniziale alla posizione `Larr-1` e con `arr[i..<s]` la sequenza dei suoi componenti nelle posizioni di `arr` dalla i alla `s-1`.

A05f.02 Cominciamo presentando le abbreviazioni più generiche e semplici.

`char` = carattere che occupa un byte, otto bits;

`int` = intero che occupa 4 bytes, 32 bits;

`seq-c` = sequenza di caratteri, ossia di bytes leggibili, oppure blank, CR ed EOF;

`seq-i` = sequenza di interi da 32 bits;

`seq-b` = sequenza di bits;

`array-char`, `array-int`;

`spis` = sequenza di puntatori a inizi di sottosequenze successive, richiede una `seq-i` e un `int` che fornisce la sua lunghezza;

`spis-c` = sequenza di caratteri organizzata in sottosequenze e relativa `spis`; richiede una `seq-i` per i puntatori, un `int` per la sua lunghezza e una `seq-c` per le sottosequenze disposte in successione;

`spis-i` = sequenza di interi organizzata in sottosequenze e relativa `spis`; richiede una `seq-i` per i puntatori, un `int` per la sua lunghezza e una `seq-i` per le sottosequenze disposte in successione;

`spis-ccc` sequenza di terne di caratteri e relativa `spis`; richieste analoghe alle precedenti.

Per i nomi e i richiami delle variabili e degli arrays collegati conveniamo quanto segue.

Se un array ha nome `arr`, alla sua lunghezza diamo il nome `arrL`.

Nomi esemplari per una `spis-c` o per una `spis-i` sono `seq` per la sequenza, `seqL` per la sua lunghezza, `seqJ[]` per la sequenza dei puntatori agli inizi delle sottosequenze ed `seqN` per il numero delle sottosequenze.

Per ciascuna di tali `spis` abbiamo `seqL = seqJ[seqN]` e la sottosequenza contraddistinta dal progressivo `j` che può assumere i valori `0, 1, ... , seqN-1` e che occupa le componenti dalla iniziale `seq[seqJ[j]]` fino all'ultima con `seqJ[j+1]-1`.

È necessario dare delle valutazioni per le estensioni che ci sembrano ragionevoli per elaborare la attuale MATeXp; queste valutazioni non vogliono essere definitive ma adattabili e le trattiamo con identificatori di variabili.

I valori massimi previsti per una lunghezza di nome `seqL` e per un numero di oggetti di nome `seqN` li identifichiamo, risp., con `seqLmx` e con `seqNmx`.

A05f.03 Un dato di ingresso presente in tutti i programmi e è la specificazione del file di governo nell'array-char `fgov[..fgovL]`; in linea di massima si prevede `fgovLmx = 24`.

Per molte manovre serve un file contenente, una in ciascuna linea, le specificazioni dei files `.tex` da elaborare, ciascuna in una linea; le specificazioni sono registrate come stringhe nullterminated nella `scSpis [[feN , feJ[<=feN] , fe[<LfeL]]`.

Previsioni: $feNmx = 180$ e $feLmx = 2000$.

I titoli dei capitoli e delle sezioni vengono trattati in memoria da due spis-c.

titoli dei capitoli: array-char $kt[<ktL]$ retto dai puntatori in $tkJ[<=kN]$; in altre parole: i titoli dei capitoli sono registrati nella spis-c $[[tkN , tkJ[<=kN] , tk[<tkL]]]$; sono previsti $tkLmx = 5000$ e $kNmx = 180$.

titoli delle sezioni nella scSpis $\langle Nsez , jts[.Ns+1] , ts[.Lts] \rangle$; previsioni: $MNsez = 1500$ e $MLts = 50000$.

collegamento fra capitoli e sezioni: $isik[ik]$ fornisce il progressivo (tipicamente identificato da is) della prima sezione del capitolo ik

sequenza delle lettere dei tomi: $cdtm[.Ntm] = "ABCDGIPTWX "$

sequenza codici capitoli: $cd[.Nkp]$, $NNkp = 200$,

$cdkp[ik] = prgr. tomo *100 + intero del suo identificativo$; e.g. il codice del capitolo G52 à 452.

A05f.04 Chiamiamo testo-M il testo sorgente di MATeXp digitalizzato in memoria.

Si distinguono caratteri, righe, (periodi,) paragrafi, sezioni e capitoli rappresentati risp. dalle lettere $c, r, (P,) p, s$ e k .

I relativi indici tipici sono $ic, ir, (iP,) ip, is, ik$.

$ic = 0, 1, \dots, Nc-1$ con $Nc \leq Mnc$

$ir = 0, 1, \dots, Nr-1$ con $Nr \leq Mnr$

$ip = 0, 1, \dots, Np-1$ con $Np \leq Mnp$

$is = 0, 1, \dots, Ns-1$ con $Ns \leq Mns$

$ik = 0, 1, \dots, Nk-1$ con $Nk \leq Mnk$

Previsione delle estensioni

capitoli Nk MNk 150 - 200

sezioni Ns MNs 1396 - 1500

paragrafi Np MNp 13972 - 15000

pagine $Npag$ $MNpag$ $180 \times 25 = 4500$

righe Nr MNr = 400 000

caratteri Nc MNc = 10 000 000

A05f.05 La sequenza di tutti i caratteri di testo-M vanno nell'array-char di base $c[<cN]$.

Elenchiamo i collegamenti tra le componenti di testo-M

lista delle coppie (paragrafo rinviate , paragrafo/sezione/capitolo rinviate nell'array int $p2psk[Np2psk]$ con $p2psk[ip] = ip + 10000*ipsk$ ove

ip = indice paragrafo rinviate $ipsk$ = indice $p/s/k$ rinviate $Np2psk$; $MNp2psk = 5000$ (ora 2856 in 158 kap.)

relazione inversa $psk2p[Np2psk]$ con $psk2p[ipsk] = ipsk + ip*10000$ array ottenuto ordinando in situ gli interi ottenuti dai $p2psk[ip]$

Vediamo il complesso degli arrays di collegamento fra le ripartizioni di testo-M, portatori dei primi caratteri di ogni ripartizione

$icir[MNr]$ $irip[MNp]$ $ipis[MNs]$ $isik[MNk]$
 $icip[NMr]$ $iris[NMs]$ $ipik[MNk]$
 $icis[NMs]$ $irik[MNk]$
 $icik[MNk]$

Inoltre si controllano i primi caratteri inutili di ogni linea con `icuir[.Nr]`

A05f.06 Si passa da un trigramma di capitolo “Tnn” al suo prgr. `ik` mediante la sequenza dei trigrammi `yk[.3][.Nk]` individuando in essa `ik` tale che `yk[1..3][ik] == "Tnn"`.

Si passa dal tetragramma di sezione “TnnL” al suo progressivo `is` trovando `ik` come sopra e cercando la posizione della sezione “L” nella `scSpis` delle lettere delle sezioni nei successivi capitoli `ys[.Ns]` ripartita per capitoli dalla `scSpis` `isik[Nk+1]`; `is` è l’unico che soddisfa la `isik[ik] <= is < isik[ik+1]` e tale che `ys[is] == "L"`.

Si passa da esagramma di paragrafo “TnnLmm” al prgr. `ip` ricavando `ik` e `is` come sopra e ottenendo `ip` considerando la sequenza dei numeri “mm” dei successivi paragrafi `yp[..Np]`, array-int ripartito per sezioni successive dalla `siSpis` `ipis[Ns+1]`: `ip` è l’unico per cui `yp[ip] == valore di "mm"` con il vincolo `ipis[is] <= ip < ipis[is+1]`.

A05 g. manovre per costruire il testo-M

A05g.01 accettazione delle ripartizioni

legge file di governo files .tex, individua sequenza file .tex da leggere e la pone nella spis-c [[feN, feJ[i=feN], fe[jfeL]]]

- ottiene seq-c dei tomi tomc[tomN] contenute con ripetizioni nei caratteri fe[feJ[kma]+2] per kma= 0..kN-1; - ottiene sequenza trigrammi dei capitoli ktg[0..2][kN] e la suddivide per capitoli con la spis-ccc [[tomN, ktgJ[<=tomN], tkg[<ktgL]

- ottiene la seq-i dei codici dei capitoli kCd[<kN], suddivisa da spis-i ktgJ[<=Nk], a partire dal relativo trigramma tkg[1:3][kma] definendo kcd[kma] = tcd[kma]*100 + ival(ktg[1:2][kma]), dove tcd[kma] è tale che ktg[0][kma]==tom[tcd[kma]].

- potendo assumere la correttezza delle sezioni e dei paragrafi dei capitoli (v.o.), scorre i files per definire sequenza dei trigrammi di sezioni e paragrafi e costruire seq-c dei caratteri delle sezioni entro i capitoli ckse[sN], anch'essa organizzata in spis-c da ktgJ[<=kN], e la seq-i degli interi che caratterizzano i paragrafi entro le sezioni ipINs[pN] organizzata nella spis-i da ipINsJ[<=sN].

- risulta stabilito il collegamento tra indici e specificazioni dei paragrafi, nonchè in seguito a semplificazione, il collegamento tra sezioni e capitoli.

```
spec(kma,sma,pma) = ktg[0:2][kma] + csec[sma] + decod(ispes[pma] e
iiicode("TnnSmm") = (kma,sma,pma),
```

dove kma = icode("Tnn"), sma = icode("S",kma), pma = decod(sma,mm);

Qui abbiamo usato functions decod, iiicode e iicode per le transcodifiche necessarie che precisereмо più avanti.

A05g.02 estrazioni da file .tex

Per elaborare il testo sorgente servono le seguenti varie operazioni di estrazione dal testo-M di suoi elementi caratterizzanti.

estrazione di rinvii interni a capitoli, a sezioni e a paragrafi;

estrazione di nidi **termine** con le relative coordinate TmmSnn;

estrazione delle occorrenze di notazioni marcate con le relative coordinate;

estrazione delle citazioni di personalità con le relative coordinate;

estrazione delle citazioni di opere con le relative coordinate;

estrazione dei rinvii coordinata, loro validazione e registrazione di coppie da paragefo a coordinata;

individuazione stringhe alfanumeriche estranee dai nidi ...,

...

e simili e verifica di assenza dal relativo elenco generale;

individuazione di ogni macro aventi forma del tipo \abcd e verifica della sua presenza nell'elenco delle macro T_EX o delle macro del redattore.

A05g.03 diagnosi dei files per il rispetto di struttura e coerenza

corre sui files forniti da un elenco predisposto nel file di governo;

corre su un file fino alla prima occorrenza di \bye;

corre su ogni linea fino a CR o prima occorrenza di “%” ;

esamina TOC iniziale della forma \csin sezioni, titoli e pagine \nop \cste ;

verifica la sequenza delle sezioni e i loro titoli;
esamina la sequenza dei paragrafi con eventuale diagnosi;
estrae titoli di capitoli e di sezioni che potrebbero essere emessi nel relativo KWIC;
estrae coordinate bersagli di rinvii \IL per controllarne la presenza nel testo
sse richiesto le registra insieme alle linee del rinvio

A05g.04 costruzione dei collegamenti entro il testo-M

costruzione dei rinvii righe di capitolo $-i$ paragrafi di capitolo
estrazione dei termini dalle linee dei files con $\%;T\ TnnSmm$ e verifica della legalità dell'esagramma;
estrazione delle linee dei files con $\%;Na\ TnnSmm$, $\%;Nc\ TnnSmm$ $\%;Nd\ TnnSmm$ $\%;Ne\ TnnSmm$ $\%;Ni\ TnnSmm$ $\%;Nm\ TnnSmm$ e verifica della legalità dell'esagramma; estrazione dalle linee dei files che iniziano con $\%;A\ TnnSmm$ e verifica della legalità dell'esagramma;
costruzione dell'indice dei termini X11;
costruzione dell'indice delle notazioni X12;
costruzione dell'indice AbbSpecVar X13;

A05g.05 estrazioni da stringhe

Molte operazioni consistono nell'esame di una stringa esprimente una riga di testo sorgente per individuare componenti qualificanti.

Oggetti che possono essere ricercati:

una data stringa NT;
una data sottostringa di un array-char determinata dalla posizione iniziale dalla posizione successiva;
nidi della forma $\{\backslash dn\ termine\}$
nidi aventi una delle forme $\backslash IL\{TnnSmm\}$, $\backslash IL\{Lnnl\}$, $\backslash It\{Tnn\}$, $\backslash IL\{:lmm\}$ o $\backslash IL\{:l\}$
nidi della forma $\backslash Ya\{it\ nome\ personalità\}$
nidi della forma $\backslash Yw\{it\ nome\ opera\}$

Sono prevedibili le ricerche di altri nidi di forme simili alle precedenti, da definire in modi che li rendano facilmente individuabili.

È anche auspicabile la possibilità di ricercare sottostringhe caratterizzate da un riconoscitore a stati finiti fornito attraverso una sua codifica.

A05g.06

Vengono studiate anche procedure aventi lo scopo di aggiornare l'intero testo-M in seguito a richieste dei tipi seguenti.

spostamenti di capitoli, sezioni o paragrafi;
eliminazioni di capitoli, sezioni o paragrafi;
inserimenti di capitoli, sezioni o paragrafi;
modifiche di elementi individuabili con algoritmi come quelli presentati in g.05.

A05 h. prospettive di sviluppi ulteriori

A05h.01 Molti motivi si possono addurre sulla opportunità di sviluppare efficaci strumenti software che consentano di supportare le attività per la pubblicistica di argomenti dell'area STRM, ossia di argomenti scientifici, tecnologici e matematici.

Un primo ordine di considerazioni riguarda la necessità di produrre tempestivamente presentazioni didattiche e divulgative riguardanti settori di ricerca e sviluppo in rapida evoluzione.

Altre considerazioni possono riguardare la necessità di presentare versioni orientate a settori applicativi specifici informazioni su sviluppi recenti che si ritiene possano influenzarli imponendo nuove tecniche, nuovi approcci e anche nuove prospettive sui piani commerciale, economico e sociale (forse anche prospettive culturali).

Queste ultime considerazioni si possono integrare con argomentazioni riguardanti la possibilità che nel prossimo futuro si abbia una crescita rapida e diffusa della domanda di conoscenze riguardanti le discipline STEM.

Si tratterebbe di un cambiamento rilevante, in particolare in Italia, stante il basso interesse attuale per questi problemi, soprattutto se rapportato all'interessamento per iniziative destinate ad esaurirsi in poco tempo e alla ormai evidente influenza decisiva delle innovazioni STEM sulla vita di tutti.

A05h.02 Un terzo ordine di considerazioni tocca la crescente disponibilità di strumenti e di infrastrutture in grado di sostenere l'efficacia delle iniziative divulgative e didattiche nella dirazione STEM.

A05h.03 Torniamo alle procedure di supporto a MATeXp e alle prospettive di ampliarne la portata e l'incisività, cercando anche di allargando lo sguardo per comprendere il possibile supporto a simili iniziative riguardanti nozioni di altri settori di studio che danno importanza alla razionalità e alle analisi quantitative.

Si possono individuare e cominciare ad esaminare due generi di prospettive.

Al primo genere appartengono le prospettive della generazione automatica di files leggibili nei linguaggi HTML, XML e WikiMedia, nonché nel formato pdf.

Al secondo genere appartengono le prospettive di generazione di files nella lingua inglese o in altre lingue che possono avere un ruolo di rilievo per la divulgazione e la didattica della matematica e le sue applicazioni.

Va anche auspicata la possibilità di elaborare stringhe per ricercare o individuare frammenti significativi, come detto in [g05], o per molti altri scopi immaginabili, servendosi di riconoscitore a stati finiti o di altre macchine sequenziali caratterizzate attraverso loro codifiche.

L'esposizione in <https://www.mi.imati.cnr.it/alberto/> e https://arm.mi.imati.cnr.it/Matexp/matexp_main.php