

Andrea Cimino

Domicilio
Via Liguria 1
56124 - Pisa (PI)

Telefono cellulare: 328-4211568
Email personale: andrea.cimino@ilc.cnr.it

Posizione attuale

Assegnista di ricerca presso l'Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC) (Centro Nazionale delle Ricerche di Pisa) nell'ambito del progetto di ricerca PRIN dal titolo "SCRIBE. Scritture brevi: semplificazione linguistica, inclusione sociale: modelli e applicazioni" sotto la supervisione del Dott. Felice Dell'Orletta. Membro del gruppo ItaliaNLP Lab presso l'ILC (<http://www.italianlp.it>).

Istruzione

Laurea Magistrale in Informatica, conseguita in data 16/12/2011 presso l'Università di Pisa, con votazione 110/110

Andrea Cimino ha completato il corso di studi con una media superiore al 28, discutendo la tesi di laurea dal titolo "*Programmazione logica con vincoli per la generazione automatica di unit-test su programmi C/C++*" (relatori Prof. Giorgio Levi e Prof. Roberto Bagnara). Nel lavoro di tesi ha affrontato la problematica di generare unit-test staticamente ed automaticamente, utilizzando tecniche di esecuzione simbolica, nel contesto dell'analizzatore ECLAIR (<http://bugseeng.com/products/eclair>). L'obiettivo della generazione automatica di unit-test è di dimostrare la raggiungibilità di porzioni di codice per alcuni valori di input del programma o, in caso contrario, dimostrarne la non raggiungibilità. A tale scopo ha ripreso la ricerca effettuata in ambito accademico (in particolare dal gruppo del Dott. Arnaud Gotlieb, Simula Research Laboratory, Oslo, Norvegia) che ha proposto come tecniche di risoluzione di questo problema tecnologie provenienti dalla programmazione logica e dalla programmazione a vincoli. Nel lavoro di tesi, inoltre, ha descritto la problematica dei comportamenti indefiniti dei linguaggi C/C++, come ad esempio l'overflow, e di come questi influenzino notevolmente il processo di generazione di test: si potrebbero infatti raggiungere porzioni di codice in seguito a computazioni non previste da un programmatore il che, in settori industriali molto delicati, non è ammissibile. Tale problematica è stata risolta facendo in modo che i vincoli generati imponessero che in nessun path di computazione fosse permessa la presenza di alcuni comportamenti indefiniti (ad esempio l'overflow). Infine ha implementato, sempre nel contesto dell'analizzatore ECLAIR, il supporto per i criteri di copertura *statement coverage* e *branch coverage*. L'implementazione del modulo di generazione di test è stata scritta in linguaggio *Prolog*.

Laurea in Informatica, conseguita in data 26/04/2006 presso l'Università degli studi di Parma, con votazione 102/110

Il lavoro tesi, dal titolo "*Un'implementazione incrementale e su aritmetica esatta del simplesso primale*", è consistito nell'integrazione nella libreria Parma Polyhedra Library di un risolutore di problemi di programmazione lineare, sotto la supervisione del relatore Dott. Enea Zaffanella e del correlatore Prof. Roberto Bagnara. La Parma Polyhedra Library (PPL, <http://www.cs.unipr.it/pp1/>) fornisce servizi fondamentali per ogni analizzatore di programmi che intenda modellare i valori numerici assunti dalle variabili del programma. Le caratteristiche principali del risolutore di problemi di programmazione lineare implementato sono due:

Utilizzo dell'aritmetica esatta: l'aritmetica *floating point* è soggetta ad approssimazioni. Sono state utilizzate delle tecniche per evitare il problema;

Incrementalità: può capitare di risolvere problemi che differiscono tra loro per pochi parametri. Un approccio incrementale migliora notevolmente le prestazioni.

La PPL è internazionalmente considerata la migliore del suo tipo: per questo motivo è stata scelta per fornire questi servizi alla GNU Compiler Collection (GCC).

Diploma di Liceo Scientifico “G.Galilei”, Lamezia Terme (CZ), con votazione 90/100

Esperienze di Ricerca

Vincitore di Assegno di Ricerca presso l’Istituto di Linguistica Computazionale “A. Zampolli” (ILC) del Consiglio Nazionale delle Ricerche (CNR) di Pisa nell’ambito del progetto di ricerca PRIN dal titolo “SCRIBE. Scritture brevi: semplificazione linguistica, inclusione sociale: modelli e applicazioni” sotto la supervisione del Dott. Felice Dell’Orletta. Selezione pubblica N.126.048/ASS. 009.2012 del 7-11-2012. (01-12-2012 - Oggi)

Andrea Cimino svolge attività di ricerca nel settore del trattamento automatico del linguaggio, basato sia su diversi metodi di apprendimento automatico supervisionati (Machine Learning), che non supervisionati. Frutto di tale attività di ricerca è stato lo sviluppo di alcuni moduli software tra i quali:

Named Entity Tagger. Componente basata su metodi di Machine Learning (“maximum entropy”, “support vector machines”), che riconosce e classifica entità nominate (Persone, Luoghi, Organizzazioni, Entità Geopolitiche) all’interno di un testo scritto in linguaggio naturale. L’accuratezza del sistema è in linea con lo stato dell’arte per l’italiano.

Native Language Identifier. Componente basata su metodi di Machine Learning che identifica la lingua nativa (L1) di una persona basandosi su una sua produzione scritta in lingua inglese (L2). Tale sistema è stato utilizzato nella partecipazione allo shared task “Native Language Identification Shared Task 2013” (<http://www.nlissharedtask2013.org>) ottenendo ottimi risultati. Gli approcci utilizzati e i risultati ottenuti sono descritti nell’articolo “*Linguistic Profiling based on General purpose Features and Native Language Identification*”.

Text-2-Knowledge (T2K). Co-autore della piattaforma linguistico-computazionale “T2K” (<http://t2k.italianlp.it>). Piattaforma sviluppata per l’estrazione di conoscenza da testi, basata su algoritmi statistici supervisionati e non, capaci di indicizzare automaticamente il testo utilizzando i termini di dominio estratti, e di creare relazioni semantiche tra le entità estratte dal testo.

Contratto per prestazione d’opera autonoma in regime di collaborazione coordinata e continuativa della durata di quattro mesi presso il Dipartimento di Matematica dell’Università di Parma (07-2006/11-2006)

Durante questa collaborazione ha implementato un risolutore di problemi di programmazione lineare intera, basato sul precedente lavoro di tesi, utilizzando la tecnica del *Branch and Bound*. Inoltre si è occupato del mantenimento della libreria (scrittura di script Automake ed Autoconf), stesura di testcase e della realizzazione di interfacce verso altri linguaggi di programmazione (Java e OCaml).

Pubblicazioni

Cimino A., Dell’Orletta F., Venturi G., Montemagni S. (2013), “Linguistic Profiling based on General-purpose Features and Native Language Identification”. In *Proceedings of Eighth Workshop on Innovative Use of NLP for Building Educational Applications, Atlanta, Georgia, June 13*, pp. 207-215.

Esperienze lavorative

Collaboratore presso KIWI s.r.l. (<http://www.kiwilocal.net/>), Pisa (06/2012-11/2012)
KIWI s.r.l. è una start-up che sta sviluppando un’applicazione per smartphone che permette di comunicare con chi si trova nelle vicinanze. In questo contesto ha lavorato sulla parte server, scritta in linguaggio Java, per aggiungere nuove funzionalità ed ottimizzare quelle già esistenti, come ad esempio la ricerca di peers in un dato raggio in modo efficiente, parte critica dell’applicazione. Si è occupato inoltre di migliorare gli algoritmi presenti sul client Android, in modo di ridurre drasticamente i consumi energetici.

Impiegato presso l'azienda Bugseng s.r.l. (<http://www.bugseng.com>), Parma (01-2011/03/2012). Si è occupato di continuare lo sviluppo del propagatore di vincoli (CSP), allo scopo di superare i limiti tecnologici imposti dal propagatore usato, il modulo Constraint Logic Programming over Finite Domains (clpfd) (<http://www.swi-prolog.org/man/clpfd.html>) di SWI-Prolog. I linguaggi principalmente usati sono stati il Prolog ed il C/C++.

Impiegato presso la ditta "Cedacri S.P.A." (<http://www.cedacri.it>) di Collecchio, Parma (02/2007-08/2009)

Nella fase iniziale della collaborazione ha programmato nei linguaggi C# e ASP.NET per lo sviluppo di applicazioni in ambito bancario. Successivamente ha lavorato all'implementazione di HP PPM (Project and Portfolio Management) (<http://www.swi-prolog.org/man/clpfd.html>) utilizzata presso "Cedacri S.P.A.". HP PPM è uno strumento che integra e unifica le attività di gestione dell'esecuzione di progetti all'interno di grandi organizzazioni: gestione della domanda, gestione finanziaria, dei tempi e delle risorse, gestione di progetti e di programmi e gestione del portfolio. Si è occupato della definizione e dell'implementazione dei processi di processi aziendali interni e di procedure Oracle PL/SQL personalizzate. Inoltre ha realizzato strumenti per la generazione automatica della reportistica, utilizzando tecnologia JSP.

Sviluppo di moduli software

Ha sviluppato numerosi componenti software. Tali componenti sono stati sviluppati in C, C++, Python, Prolog, Java. Segue un elenco di alcuni dei vari componenti sviluppati nel tempo, tale elenco non include le numerose personalizzazioni sviluppate:

Sviluppo di un risolutore di problemi di programmazione lineare e programmazione lineare intera all'interno della libreria *Parma Polyhedra Library*, sviluppato presso il Dipartimento di Matematica dell'Università di Parma. Linguaggi utilizzati: C/C++

Realizzazione di interfacce OCaml e Java per la libreria *Parma Polyhedra Library*, al fine di poter utilizzare la libreria in linguaggi diversi dal C++, sviluppato presso il Dipartimento di Matematica dell'Università di Parma. Linguaggi utilizzati: OCaml, Java, C.

Sviluppo di applicazioni in ambito bancario. Linguaggi e tecnologie utilizzate: C#, .Net, SQLServer, Oracle, Java, JSP.

Sviluppo di un generatore di vincoli CSP in grado di catturare la semantica del C/C++ per la generazione automatica di test. Lavoro svolto durante la tesi di laurea magistrale. Linguaggio utilizzato: Prolog.

Sviluppo di un propagatore di vincoli CSP, sviluppato presso l'azienda Bugseng s.r.l. di Parma. Linguaggio utilizzato: Prolog.

Sviluppo dell'applicazione "Kiwi-local", (<http://www.kiwilocal.net/>) sia in ambito server che client, presso l'azienda Kiwi s.r.l. di Pisa. Linguaggi, tecnologie e piattaforme utilizzate: Java, Restlet, Python, Android.

Sviluppo di moduli software per l'analisi del linguaggio naturale descritti nel paragrafo "Esperienze di Ricerca". Linguaggi e tecnologie utilizzate: Python, Django, libsvm (modulo per lo sviluppo basati sulle support vector machines), maxent (modulo per lo sviluppo basati sulla maximum entropy), sigmajs (modulo per la visualizzazione di grafi).

Riconoscimenti

Corso di Natural Language Processing @ Coursera (02/2013-05/2013)

Andrea Cimino ha sostenuto il corso di Natural Language Processing (Prof. Michael Collins) sulla piattaforma online Coursera e ha superato l'esame con giudizio "with distinction". Tale corso ha trattato numerosi argomenti legati al trattamento del linguaggio naturale, tra i quali: modelli del linguaggio, modelli di Markov nascosti (Hidden Markov Models, HMM), grammatiche probabilistiche libere da contesto (probabilistic context-free

grammars, PCFG), modelli log-lineari (log-linear models), apprendimento non supervisionato e semisupervisionato, il problema dell'analisi sintattica (syntactic parsing).

Link alla pagina del corso: <https://www.coursera.org/course/nlangp>

Vincitore del bando Spinner 2013 (01/2011-03/2012)

La Sovvenzione Globale Spinner 2013 (<http://www.spinner.it/>) è un programma della regione Emilia Romagna che permette di sviluppare progetti di ricerca, sviluppo precompetitivo e trasferimento tecnologico dal mondo dell'Università a quello dell'industria. L'azienda con la quale ha collaborato nel contesto di questo programma è stata "Bugseng s.r.l." di Parma.

Lingue conosciute

Madrelingua: italiano e polacco;

Ottima conoscenza della lingua inglese scritta e parlata.

Linguaggi di programmazione conosciuti

C, C++, C#, Python, Prolog, Java.

Tecnologie web conosciute

Django, JSP, Restlet, Javascript, PHP.

Librerie per lo sviluppo di algoritmi basati sul machine learning

libsvm (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>)

maxent (<http://homepages.inf.ed.ac.uk/lzhang10/maxent.html>)

Luogo e data

Pisa, 27/08/2013

Firma

