

Università degli Studi della Calabria

Dipartimento di Matematica

Dottorato di Ricerca in Matematica ed Informatica

XX Ciclo

Settore Disciplinare INF/01 INFORMATICA

Tesi di Dottorato

**Efficient Evaluation of
Disjunctive Logic Programs**

Gelsomina Catalano

Supervisor

Prof. Nicola Leone

Dott.ssa Simona Perri

Coordinatore

Prof. Nicola Leone

Anno Accademico 2007 - 2008

Efficient Evaluation of Disjunctive Logic Programs

Gelsomina Catalano

*Dipartimento di Matematica,
Università della Calabria
87036 Rende, Italy
email : catalano@mat.unical.it*

Sommario

Agli inizi degli anni '80, Jack Minker propose di accrescere la potenza della programmazione logica consentendo l'utilizzo della disgiunzione nelle teste delle regole e specificando come l'assunzione di mondo chiuso potesse essere estesa al linguaggio risultante, chiamato Programmazione Logica Disgiuntiva (DLP) [Minker, 1982; 1994]. Più tardi, Michael Gelfond e Vladimir Lifschitz fornirono una semantica per la DLP, detta Answer Set Semantics [Gelfond and Lifschitz, 1991], che ha ricevuto larghi consensi nella comunità scientifica ed è ora generalmente adottata per la DLP (detta anche Answer Set Programming – ASP). In accordo a tale semantica un programma logico disgiuntivo può avere più modelli alternativi (ma anche nessuno) ognuno corrispondente a una possibile visione del mondo rappresentato dal programma.

Il linguaggio per la rappresentazione della conoscenza DLP è molto espressivo in un senso matematicamente preciso; la DLP può rappresentare *ogni* problema nella classe di complessità Σ_2^P (NP^{NP}) [Eiter *et al.*, 1997b]. Dunque, sotto assunzioni ampiamente accettate, la DLP risulta strettamente più espressiva della programmazione logica *normale* (*senza disgiunzione*), la cui espressività è limitata alle proprietà decidibili in NP. L'espressività della Programmazione Logica Disgiuntiva, ha importanti implicazioni pratiche, poichè esistono problemi che possono essere rappresentati tramite un programma logico disgiuntivo, ma che non è possibile esprimere con programmi logici senza disgiunzione, considerata la loro complessità [Eiter *et al.*, 1997b]. Inoltre, la disgiunzione consente di rappresentare in modo più semplice e naturale problemi appartenenti a classi di complessità più basse. La DLP, con la Answer Set Semantics, è oggi ampiamente riconosciuta come uno strumento potente per la rappresentazione della conoscenza e il ragionamento di senso comune [Baral and Gelfond, 1994; Lobo *et al.*, 1992; Wolfinger, 1994; Eiter *et al.*, 1999; Gelfond and Lifschitz, 1991; Lifschitz, 1996; Minker, 1994; Baral, 2002].

L'elevata complessità della DLP ha scoraggiato per molti anni la realizzazione di sistemi che implementassero tutte le caratteristiche di tale linguaggio. Dopo alcuni anni di ricerca sia teorica che algoritmica, oggi esistono diversi sistemi che supportano la DLP o parte di essa. Oltre ai sistemi di programmazione logica (non disgiuntivi) Smodels [Simons *et al.*, 2002] e ASSAT [Lin and Zhao, 2002], sono disponibili anche alcuni sistemi di programmazione logica (disgiuntiva): DLV [Leone *et al.*, 2006], GnT [Janhunen *et al.*, 2003], and cmodels-3 [Lierler, 2005].

In questa tesi ci concentriamo sul sistema DLV, che è riconosciuto essere lo stato dell'arte della Programmazione Logica Disgiuntiva. DLV è ampiamente sfruttato in tutto il mondo sia a scopo di ricerca che didattico. Per esempio, è stato impiegato al CERN, il Laboratorio Europeo di Fisica delle Particelle di Ginevra, per un'applicazione di Basi di Dati deduttive che coinvolge la manipolazione di conoscenza complessa su basi di dati di grandi dimensioni. La compagnia polacca Rodan Systems S.A. sfrutta DLV in uno strumento per scoprire le manipolazioni dei prezzi e l'uso non autorizzato di informazioni confidenziali. Noi crediamo che la forza di DLV – la sua espressività e l'implementazione solida – lo renda attrattivo per tali applicazioni complesse. Anche dal punto di vista dell'efficienza, esso è competitivo con i più avanzati sistemi in quest'area come confermano i recenti confronti e valutazione delle prestazioni [Leone *et al.*, 2006; Dix *et al.*, 2003; Arieli *et al.*, 2004], e i risultati della Prima Competizione di Sistemi Answer Set Programming <http://asparagus.cs.uni-potsdam.de/contest/>, in cui DLV è risultato vincitore per le categorie DLP e MGS Competition (detta anche categoria “Royal”). Lo sviluppo di DLV è iniziato nel 1996 al Politecnico di Vienna, nell'ambito di un progetto finanziato dalla Austrian Science Funds (FWF); oggi, DLV è oggetto di una cooperazione internazionale tra l'Università della Calabria e il Politecnico di Vienna.

Il presente lavoro di tesi è incentrato sullo studio della Programmazione Logica Disgiuntiva e l'ottimizzazione del sistema DLV, che implementa la DLP.

I nostri studi hanno evidenziato che negli ultimi anni, la disponibilità di sistemi DLP affidabili, ha indotto a sfruttare la DLP in diverse aree applicative, ma i sistemi attuali non sono sufficientemente efficienti per molte di queste applicazioni. Il nostro lavoro affronta tale aspetto, proponendosi di superare questa limitazione, migliorando l'efficienza dei sistemi DLP e del sistema DLV tramite il progetto e l'implementazione di nuove tecniche di ottimizzazione.

I moduli della maggior parte dei sistemi DLP operano su un'istanziamento ground del programma in input, cioè un programma che non contiene alcuna variabile, ma è semanticamente equivalente all'input originale [Eiter *et al.*, 1997c]. Ogni programma P in input, è inizialmente sottoposto alla cosiddetta procedura di istanziamento (detta anche istanziatore) che calcola, a partire da P , un programma ground P' semanticamente equivalente. Poiché questa fase può essere molto costosa, avere un buon istanziatore è un aspetto cruciale dei sistemi DLP. La ragione è dovuta al fatto che ogni atomo di ciascuna regola può essere istanziato utilizzando ogni costante dell'Universo di Herbrand del programma, con una evidente

esplosione esponenziale. L'istanziatore dovrebbe essere in grado di produrre un programma ground P' avente gli stessi answer set di P e tale che: (i) P' sia calcolato efficientemente da P , e (ii) P' sia il più piccolo possibile, e quindi possa essere valutato più efficientemente da un solver DLP.

Alcune applicazioni della DLP in aree emergenti come la gestione della conoscenza e l'integrazione delle informazioni, in cui devono essere processate grandi quantità di dati, hanno reso evidente la necessità di migliorare significativamente gli istanziatori DLP.

La nostra attenzione è stata rivolta al modulo di istanziazione di DLV, investigando nuove possibili direzioni per aumentarne l'efficienza. In particolare, in questa tesi, presentiamo due proposte per migliorare la procedura di istanziazione:

- Una nuova tecnica di *Backjumping* per l'istanziatore di DLV e
- Nuove tecniche di *Indicizzazione* per l'istanziatore di DLV

Di seguito descriviamo brevemente queste due linee di ricerca.

Backjumping. Proponiamo di sfruttare tecniche di backjumping nel processo di istanziazione di una regola di DLV. Progettiamo un algoritmo basato su una struttura di backjumping che riduce la taglia del programma ground generato e ottimizza il tempo di esecuzione necessario a produrlo. In particolare, data una regola r che deve essere resa ground, tale algoritmo sfrutta informazioni semantiche e strutturali su r , per calcolare efficientemente le istanze ground di r , evitando la generazione di regole "inutili". Cioè per ogni regola r si calcola solo un sottoinsieme rilevante delle sue istanze ground, preservandone la semantica.

Implementiamo questo algoritmo in DLV e conduciamo un'attività di sperimentazione su un'ampia collezione di problemi. I risultati sperimentali sono molto positivi: la nuova tecnica migliora sensibilmente l'efficienza del sistema DLV su molte classi di problemi.

Indicizzazione. Proponiamo di adoperare tecniche di indicizzazione (in memoria primaria) per migliorare le performance della procedura di istanziazione di DLV, cioè tecniche per il progetto e l'implementazione di strutture dati che permettano di accedere più efficientemente a grandi datasets. In particolare, adattiamo al nostro contesto, una tecnica classica di indicizzazione sul primo argomento e proponiamo una strategia di indicizzazione "on demand" in base alla quale gli indici

non sono predeterminati, ma piuttosto vengono calcolati su un argomento qualsiasi durante la valutazione (e solo se sfruttabili). In più definiamo due euristiche che possono essere usate per stabilire l'argomento più appropriato da indicizzare, quando esistono diverse possibilità.

Inoltre, implementiamo le tecniche di indicizzazione proposte in DLV e confrontiamo sperimentalmente le nostre strategie su una collezione di problemi provenienti da diversi domini comprese anche istanze di problemi reali. Il quadro generale risultante dagli esperimenti è molto positivo:

- Tutte le tecniche proposte e testate permettono di ottenere notevoli miglioramenti per l'esecuzione dell'istanziamento.
- Lo schema di indicizzazione on demand dà risultati migliori rispetto al classico schema sul primo argomento in un numero maggiore di casi e le performance migliorano particolarmente quando viene utilizzata una buona euristica.

In definitiva, i metodi proposti migliorano sensibilmente l'efficienza dell'istanziatore di DLV, consentendo l'utilizzo del sistema anche in applicazioni data-intensive. Comunque, per verificare ulteriormente la potenza del nuovo istanziatore conduciamo una profonda analisi sperimentale per confrontarlo con gli altri due più popolari istanziatori, Lparse [Niemelä and Simons, 1997; Syrjänen, 2002] e *GrinGo* [Gebser *et al.*, 2007b]. L'analisi conferma che, il nuovo istanziatore ha performance migliori degli altri su tutti i problemi testati, mentre il vecchio mostra performance simili agli altri.

I risultati presentati in questa tesi sono rilevanti anche per altri due aspetti: da una parte, l'istanziatore di DLV può essere sfruttato proficuamente da altri sistemi che non hanno un istanziatore proprio, per esempio ASSAT [Lin and Zhao, 2002] e Cmodels [Lierler and Maratea, 2004; Babovich, since 2002]. Infatti, questi sistemi possono usare DLV per ottenere il programma ground (lanciando DLV con l'opzione "-instantiate"), e poi applicare le proprie procedure per la valutazione del programma ground; d'altra parte, tutti i metodi proposti sono abbastanza generali e quindi, possono essere facilmente adattati per essere integrati nella fase di calcolo di altri istanziatori. In realtà la nostra tecnica di backjumping è stata già integrata in altri due istanziatori, *GrinGo* e FO+ [Wittocx *et al.*, 2008], con buoni risultati. Le buone performance di *GrinGo* in alcuni degli esperimenti condotti, infatti, sono dovute all'utilizzo della nostra tecnica.

I principali contributi della tesi possono essere riassunti come segue:

1. Studiamo la DLP, la sua complessità e il suo utilizzo per la rappresentazione della conoscenza e il ragionamento non monotono.
2. Progettiamo un nuovo metodo, basato su una tecnica di backjumping, che consente di ridurre sia la taglia dell'istanziamento dei programmi DLP che il tempo necessario per generarla. Implementiamo il metodo proposto in DLV e conduciamo un'attività di sperimentazione.
3. Definiamo due nuove strategie di indicizzazione, per ottimizzare il tempo di istanziazione di DLV. Inoltre, implementiamo le tecniche proposte nel sistema DLV ed effettuiamo un'analisi sperimentale.
4. Confrontiamo l'istanziatore di DLV con altri due istanziatori, Lparse e *GrinGo* e discutiamo i risultati.