



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

## استدلال کننده Hermit OWL

چکیده :

Hermit را می توان تنها استدلال کننده ای نام برد که ما می دانیم به طور کامل از استاندارد OWL2 پشتیبانی می کند و به طور صحیح در مورد ویژگی ها و دسته ها استدلال می کند . این استدلال کننده بر اساس حساب جدید hypertableau می باشد که به عملکرد مسایل به دلیل عدم جبرگرایی و اندازه مدل ( منابع اولیه پیچیدگی در استدلال کننده های OWL با جدی ترین فناوری ) می پردازد . همچنین Hermit تعدادی از بهینه سازی های بدیع از جمله رویه دسته بندی هستی شناسی بهینه را یکپارچه می سازد . تست های ما نشان می دهند که Hermit در مقایسه با استدلال کننده های موجود جدولی به خوبی اجراء می کند و اغلب در زمانی بسیار سریع می باشد که هستی شناسی های پیچیده را دسته بندی می کند .

### 1- مقدمه

Hermit یک سیستم استدلال OWL مبتنی بر حساب جدید hypertableau می باشد [12] . Hermit همانند سیستم های موجود مبتنی بر جدول می تواند کل وظایف استدلال را برای تست رضایت بخش هستی شناسی کاهش می دهد و عدم رضایت بخشی هستی شناسی را از طریق تلاش برای ساخت مدل مناسب ثابت می کند . تکنیک جدل هایپر وقتی با محاسبات جدول مقایسه شده بود در واقع نشان داده است می تواند تا حد زیادی اندازه مدل های ساخت یافته و حدس غیر قطعی بکار رفته برای کشف کل ساختار های ممکن را کاهش دهد . علاوه بر این ، Hermit از الگوریتم دسته بندی جدید بهره می برد که تا حد زیادی تعداد رده بندی و تست های مورد نیز برای دسته بندی هستی شناسی معین را کاهش می دهد .

تست های ما نشان می دهند که Hermit در زمان دسته بندی هستی شناسی های با فرآیند نسبتا ساده سریع تر از دیگر استدلال کننده های OWL می باشد و معمولا در زمانی که هستی شناسی های دشوار تر دسته بندی می شوند بسیار سریع تر است . علاوه براین ، اخیرا Hermit تنها استدلال کننده شناخته شده برای ما می باشد که از استاندارد

OWL2 به طور کامل پشتیبانی می کند : این استدلال کننده از کل انواع داده مشخص شده در استاندارد پشتیبانی می کند و به طور صحیح در مورد ویژگی ها و همچنین در مورد دسته ها استدلال می کند . اکثر دیگر استدلال کننده ها تنها از مجموعه فرعی نوع داده های OWL2 پشتیبانی می کنند [11] و کل دیگر استدلال کننده های OWL شناخته شده برای ما تنها استدلال مبتنی بر نحو را در زمانی که ویژگی ها را دسته بندی می کنیم ، اجراء می کنند و از اینرو ممکن است در کشف روابط غیر بدیهی اما از نظر نحوی با ویژگی فرعی فراهم شده موفق نباشند [4] .

همچنین Hermit شامل کارکرد پذیری غیر استاندارد می باشد که اخیرا در هر نوع سیستم دیگر در دسترس نمی باشد . بویژه ، Hermit از استدلال با هستی شناسی های حاوی گراف های توصیف حمایت می کند . همانطور که در رفرنس [10] نشان داده شده است ، گراف های توصیف در واقع نمایش اشیای ساخت یافته یعنی اشیای متشکل از چندین بخش متصل به هم در روش های اختیاری را توجیه می کنند . ای اشیاء در هستی شناسی های بیو پزشکی نظیر FMA و GALEN فراوان هستند اما نمی توانند به طور صادقانه در OWL نشان داده شوند .

Hermit به صورت یک کتابخانه جاوا منبع باز در دسترس می باشد و شامل API جاوا و واسط خط فرمان ساده می باشد . ما از OWL API [6] هم به عنوان بخشی از واسط جاوا عمومی و هم به عنوان تجزیه کننده برای فایل های OWL استفاده می کنیم ؛ از اینرو Hermit می تواند هستی شناسی ها را در هر فرمتی که از طریق OWL API از جمله EDF/XML ، ترکیب کارکردی OWL ، KSS ، OBO و هندل شده اند ، پردازش نماید .

## 2 - معماری و بهینه سازی ها

O را در هستی شناسی OWL می توان به سه بخش تقسیم نمود : اصول بدیهی دارایی ، اصول بدیهی دسته و سند ها . اینها با RBox R ، TBox T و ABox A پایگاه دانش منطق توصیف  $K=(R,T,A)$  شباهت دارند . کل وظایف استدلال اساسی از جمله رده بندی بررسی را می توان برای تست رضایت مندی چنین پایگاه دانشی کاهش داد . برای مثال ،  $\mathcal{K} \models A \subseteq B \text{ iff } (R, T, A \cup \{A \sqcap \neg B(s)\})$  رضایت بخش نمی باشد و S در آنجا یک فرد تازه می باشد ( یعنی فردی که در k رخ نمی دهد ) .

برای این نشان دهیم یک پایگاه دانش  $k=(R,T,A)$  رضایت بخش است ، الگوریتم جدول یک اقتباس را می سازد - توالی  $A_0, A_1, \dots, A_n$  ABoxes که در آنجا  $A_0=A$  می باشد و  $A_i$  از  $A_{i-1}$  از طریق کاربرد یک قاعده واسط بدست

می آید. قواعد استنباط باعث فهمیدن اطلاعات در اصول بدیهی R و صریح شدن T می شوند و از اینرو ABox A را به سمت مدل k استنتاج می کنند. الگوریتم یا در صورتی خاتمه می یابد که قاعده استنباط برای An قابل اعمال نمی باشد که هر An در آن معرف مدل k می باشد یا اگر An حاوی تضاد آشکار باشد که ساختار مدل در آن موفق نبوده است. قواعد استنباط زیر به طور معمول در محاسبات جدول DL استفاده می شوند.

- قاعده  $\sqcup$ : با در نظر گرفتن  $(C_1 \sqcup C_2)(s)$ ، یا از  $C_1(s)$  یا  $C_2(s)$  اشتقاق گرفته می شود.

- قاعده  $\sqcap$ : با در نظر گرفتن  $(C_1 \sqcap C_2)(s)$ ، از  $C_1(s)$  و  $C_2(s)$  استخراج می گردد.

- قاعده  $\exists$ : با در نظر گرفتن  $(\exists R.C)(s)$ ،  $R(s,t)$  و  $C(t)$  را برای t یک فرد تازه را استخراج می کند.

- قاعده  $\forall$ : با در نظر گرفتن  $(\forall R.C)(s)$  و  $R(s,t)$ ،  $C(t)$  را استخراج می کند.

- قاعده  $\sqsubseteq^-$ : با در نظر گرفتن امر بدیهی  $C \sqsubseteq D$  و یک فرد s،  $(\neg C \sqcup D)(s)$  را استنتاج می کند.

قاعده  $\sqcup$  غیر قطعی می باشد و پایگاه دانش k تنها و تنها در صورتی غیر رضایت بخش است که کل گزینه ها به تضاد منجر می شوند.

یا-شاخه بندی: این تکنیک مبتنی بر مورد برای رسیدگی به تفکیک در بعضی مواقع یا-شاخه بندی نامیده می شود.

قاعده  $\sqsubseteq$  یک منبع اصلی برای یا-شاخه بندی می باشد چون یک تفکیک را برای هر امر بدیهی TBox به هر فرد در

یک ABox می افزاید حتی اگر امر بدیهی مشابه معادله با بند Horn می باشد و از اینرو به طور ذاتی غیر قطعی می

باشد. این چنین کاربرد بدون تبعیض قاعده  $\sqsubseteq$  می تواند یک منبع مهم ناکارآمدی باشد. و این موضوع از طریق انواع

بهینه سازی های جذب از جمله جذب نقش [14] و جذب دوتایی [8] رسیدگی شده است.

الگوریتم فرآیند Hermit از طریق بازنویسی منطق توصیف بر طبق شکلی به این بهینه سازی تعمیم داده می شود

که به کل این قبیل جذب ها اجازه می دهد که قرار است به طور همزمان اجراء شوند و همچنین اجازه می دهد تا انواع

اضافی جذب در حساب های جدول استاندارد غیر ممکن شوند. علاوه براین، Hermit به طور واقعی مفاهیم DL را

برای کاهش بیشتر غیر قطعی بازنویسی می کند و از اینرو قادر است تا بهینه سازی های شیوه جذب را از به طور نافذ تر

اعمال نماید.

و - شاخه بندی : معرفی افراد جدید در قاعده  $\exists$  را در بعضی مواقع و- شاخه بندی می نامند و این یک منبع مهم دیگر برای ناکارآمدی در الگوریتم های جدول می باشد [2]. الگوریتم های جدول برای تضمین خاتمه از بلوک بندی استفاده می کنند تا از کاربرد تکراری قاعده  $\exists$  به طور نامحدود جلوگیری بعمل آورند [7]. بلوک بندی استاندارد تنها در امتداد شاخه تکی افراد تازه اعمال می گردد. Hermit از راهبرد بلوک بندی هرکجا تهاجمی استفاده می کند [12] که می تواند اندازه مدل های تولید شده را از طریق فاکتور نمایی کاهش می دهد و این کار به طور اساسی باعث بهبود عملکرد واقعی در بعضی هستی شناسی های دشوار و پیچیده می گردد.

همچنین Hermit تلاش می کند تا اندازه مدل تولید شده را با استفاده از تکنیکی کاهش دهد که استفاده مجدد انفرادی نامیده شد : وقتی یک R.C<sup>3</sup> وجودی در حال توسعه می باشد ، در ابتدا تلاش می کند تا از بعضی افراد موجود با برچسب C برای ساخت مدل استفاده مجدد نماید و تنها اگر این ساختار مدل موفق نباشد به کاهش یک فرد جدید مبادرت می ورزد . این رویکرد به Hermit اجازه می دهد تا مدل های بدون شکل درخت را مورد بررسی قرار دهد و اندازه مدل های تولید شده را برای هستی شناسی هایی به طور چشمگیر کاهش می دهد که ساختار های پیچیده نظیر هستی شناسی های اناتومی را توضیح می دهند . از اینرو افرادی که از نو استفاده شده اند به طور معنایی با مفاهیم اسنی برابر هستند و از اینرو بهره های عملکرد به دلیل استفاده مجدد انفرادی بی نهایت به هندلینگ کافی اسمی ها وابسته هستند . از اینرو Hermit از قاعده معرفی اسمی استفاده می کند که عدم جبرگرایی را کاهش می دهد و در معرفی اسم های جدید اش محافظه کار تر می باشد .

### پنهان کردن برچسب های بلوک زنی

هر جایی که بلوک زنی از ساختار مدل تکراری در مسیر تست امکان رضایت مندی تکی اجتناب می کند . Hermit بیشتر این رویکرد را توسعه می دهد تا از ساختار تکراری در میان کل مجموعه تست های رضایت مندی دوری گزیند . Hermit به طور مفهومی بجای اجرای تست های متفاوت n از طریق ساخت مدل های مختلف n یک تست تکی را اجراء می کند که یک مدل تکی حاوی تکه های مستقل n را می سازد . هر چند دو تکه به هم وصل نمی شوند ، افراد در یک تکه می توانند آنها را در تکه دیگر بلوک نمایند و تا حد زیادی اندازه مدل ترکیب شده را کاهش میدهند . در عوض ، نمایش فشرده مدل تولید شده بعد از هر تست با هدف بلوک زنی در تست های آینده حفظ می گردد . این

راهبرد بی دقت با هستی شناسی های حاوی غیر واقعی ها سازگار نمی باشد که می توانست مدل ها را از تست های مستقل متصل نماید .

این بهینه سازی برای دستیابی به نتایجی کلیدی بوده است که ما در بخش سوم معرفی می کنیم . برای مثال ، تنها یک تست رضایتمندی در GALEN پر هزینه است چون بخش اساسی مدل TBOX را محاسبه می کند ؛ کل تست های رضایتمندی بعدی از بخش های بزرگ مدل از نو استفاده می کنند .

**بهینه سازی های دسته بندی :** الگوریتم های استدلال DL اغلب در عمل استفاده می شوند تا دسته بندی پایگاه دانش  $K$  را محاسبه نمایند - یعنی تعیین می کند آیا  $K \models A \subseteq B$  برای هر جفت مفاهیم اتمی  $A$  و  $B$  در  $K$  رخ می دهد یا خیر . به طور واضح ، الگوریتم دسته بندی غیر دقیق شامل عدد درجه دوم تست های رده بندی خواهند بود که هر یک از آنها می توانند به طور بالقوه گران باشند . انواع بهینه سازی برای دستیابی به سطوح قابل قبول عملکرد توسعه یافته اند که تعداد تست های رده بندی [3] و زمان مورد نیاز برای هر تست را کاهش می دهند .

Hermit از تکنیک دسته بندی جدید استفاده می کند و از ویژگی های منحصر به فرد حساب دیفرانسیل جدید سیستم برای بهینه سازی بیشتر تکنیک بهره برداری می نماید . طور خاص وقتی Hermit تلاش می کند تا مدل ارا از  $K \cup \{A(a)\}$  بسازد ، همچنین قادر است تا از اطلاعات در  $A$  بهره برداری نماید تا مقدار زیاد اطلاعات را در مورد رده بندی کننده و غیر رده بندی کننده های  $A$  استنتاج نماید ، اطلاعاتی که می توانند به طور کارآمد از طریق رویه دسته بندی جدید بهره برداری شوند [4] .

جدول 1 : آمار های آنتولوژی ها

Ontology Name	بیانگری DL	دسته ها asses	Properties	TBox	RBox
EMap (Feb09)		3737	2	13730	0
GO Term DB (Feb06)	$\mathcal{EL}++$	20526	1	28997	1
DLP ExtDnS 397	$SHIN$	96	186	232	675
LUBM (one university)	$AL\mathcal{E}HI^+(D)$	43	32	142	51
Biological Process (Feb09)	$\mathcal{EL}++$	16303	5	32286	3
MGED Ontology	$ALCOF(D)$	229	104	452	21
RNA With Individual (Dec09)	$SRIQ(D)$	244	93	364	310
NCI Thesaurus (Feb09)	$ALCH(D)$	70576	189	100304	290
OBI (Mar10)	$SHOIN(D)$	2638	83	9747	150
FMA Lite (Feb09)	$AL\mathcal{E}I^+$	75145	3	119558	3
FMA-constitutional part (Feb06)	$ALCOIF(D)$	41648	168	122695	395
GALEN-doctored	$AL\mathcal{E}HIF^+$	2748	413	3937	799
GALEN-undoctored	$AL\mathcal{E}HIF^+$	2748	413	4179	800
GALEN-module1	$AL\mathcal{E}HIF^+$	6362	162	14515	219
GALEN-full	$AL\mathcal{E}HIF^+$	23136	950	35531	2165

### 3 نتایج تجربی

ما برای ارزیابی الگوریتم استدلال خودمان در عمل Hermit را با استدلال کننده های جدول با جدید ترین فناوری ما برای ارزیابی الگوریتم استدلال خودمان در عمل Hermit و FaCT++ 1.5.3 و Pellet 2.3.0 [13] مقایسه کرده ایم [15].

ما تعداد هستی شناسی های تست استاندارد را انتخاب کرده ایم و زمان مورد نیاز را برای دسته بندی آنها با استفاده از هر یک استدلال کننده های ذکر شده اندازه گیری کرده اند. Hermit بر خلاف Pellet و FaCT++ شامل استدلال کننده تخصیص یافته برای هر تکه کنترل شدنی OWL 2 می باشد. از اینرو، ما عمدتاً بر هستی شناسی هایی تمرکز می کنیم که از اکثر یا همه قدرت بیانی در دسترس در OWL 2 استفاده می کنیم. کل تست های در 2.7 GHz MacBook Pro با 8 گیگا بیت حافظه فیزیکی اجراء گردیده اند. اگر دسته بندی کل حافظه در دسترس را تخلیه نمایند یا اگر فراتر از تایموت 20 دقیقه ای باشد، تلاش دسته بندی موفق نمی گردد.

بخش عمده هستی شناسی های تست از طریق کل سه استدلال کننده در سریع ترین زمان ممکن دسته بندی شده بودند. عملکرد Hermit برای این هستی شناسی های بدیهی با دیگر استدلال کننده ها قابل مقایسه بوده است. از اینرو، ما در اینجا تنها نتایج تست را برای هستی شناسی های جالب بررسی می کنیم یعنی هستی شناسی هایی که یا بدهی نیستند یا استدلال کننده های تست شده در آن یک تفاوت مهم را در عملکرد نشان داده اند.

نتایج تست های ما در این هستی شناسی هایی جالب در جدول دو خلاصه می شوند . نظر به این که Hermit دارای هندلینگ خاص برای تکه های رام شدنی OWL 2 نبوده است ، عملکرد Hermit در چنین هستی شناسی هایی ممکن نیست رقابتی باشد . برای مثال ، FaCT++ یک مزیت را در زمانی نشان می دهد که دسته بندی هستی شناسی هایی که در تکه نحوی از قبل تعریف شده واقع می شوند که برای ان از تکنیک استدلال کارآمد تر استفاده می کند [16] . نسخه های متفاوت GALEN به طور معمول برای تست عملکرد استدلال کننده های DL استفاده شده اند . نسخه کامل هستی شناسی ( که GALEN- کامل نامیده شده است ) را نمی توان از طریق هر نوع استدلال کننده ها پردازش نمود . از اینرو ، ما ماژول را بر اساس مفهوم تکی GALEN- کامل با استفاده از تکنیک رفرنس [ 5 ] با هدف تعیین این سؤال استخراج کرده این که آیا تکنیک های مدولاسیون ممکن است دسته بندی ها را امکان پذیر سازند یا خیر . از اینرو ، هر چند ماژول بسیار کوچک تر از هستی شناسی مامل می باشد هیچ استدلال کننده ای قادر نبوده است تا ان را دسته بندی نماید . تحلیل ما نشان داده است که استدلال کننده ها به دلیل تعداد زیاد امر های بدیهی دوره ای یک ABoxes بی نهایت بزرگ می سازند و در نهایت کل حافظه در دسترس را تخلیه می کنند . بخش اساسی FMA یک سری مشخصه های مشابه را نمایش می دهد اما این با توسعه کمتر است و Hermit و FaCT++ قادر بودند تا ان را دسته بندی نمایند / نسخه های ساده شده متنوع GALEN به دلیل شکست استدلال کننده های DL برای پردازش نسخه های ساده شده متنوع GALEN اغلب در عمل استفاده شده اند . همانطور که جدول 2 نشان می دهد ، این هستی شناسی ها می توانند برای استدلال کننده های با جدید ترین فناوری چالش برانگیز باشند . از اینرو ، Hermit می تواند آنها را تا حدی به طور کارآمد دسته بندی نماید .

جدول 2 : نتایج ارزیابی عملکرد



Ontology Name	Classification Times (seconds)		
	Hermit	Pellet	FaCT++
EMap (Feb09)	1.1	0.4	34.2
GO Term DB (Feb06)	1.3	1.3	6.1
DLP ExtDnS 397	1.3	timeout	0.05
LUBM (one university)	1.7	0.7	152.7
Biological Process (Feb09)	1.8	4.0	8.0
MGED Ontology	2.1	19.6	0.04
RNA With Individual (Dec09)	2.7	0.8	102.9
NCI Thesaurus (Feb09)	58.2	12.3	4.4
OBI (Mar10)	150.0	timeout	17.2
FMA Lite (Feb09)	211.1	timeout	timeout
FMA-constitutional part (Feb06)	1638.3	timeout	396.9
GALEN-doctored	1.8	timeout	2.5
GALEN-undoctored	6.7	out of mem.	11.6
GALEN-module1	out of mem.	timeout	timeout
GALEN-full	out of mem.	timeout	timeout

#### 4- نتیجه گیری ها و مسیر های آینده

ما Hermit را توضیح داده ایم که یک استدلال کننده مبتنی بر الگوریتم ها و بهینه سازی های جدید می باشد . Hermit به طور کامل از استاندارد OWL2 پشتیبانی می کند و مزیت های عملکرد مهم را نسبت به دیگر استدلال کننده ها در سراسر دامنه گسترده هستی شناسی های جهان واقعی اظهاری نشان می دهد . هر چند Hermit همیشه سریع ترین نمی باشد ولی عملکرد نسبتا قدرتمند را در هستی شناسی های تست شده ما نمایش می دهد و همانطور که در نتایج امان نشان داده شده است ، هرگز برای دسته بندی یک هستی شناسی در بدنه تست موفق نبوده است که به طور موفقیت آمیزی از طریق یکی از استدلال کننده های دیگر رسیدگی نشده است . همچنین Hermit شامل پشتیبانی از بعضی خصیصه های هستی شناسی غیر استاندارد نظیر گراف های توصیف می باشد .

ما در حال تداوم توسعه Hermit هستیم و تکنیک های بهینه سازی اصلاح شده و جدید را کشف می نماییم . همچنین ما به توسعه کارکرد پذیری اش ادامه می دهیم : برای مثال جدید ترین نسخه از زبان جستجو SPARQL 1.1 پشتیبانی می کند [1] . همچنین ما در حال بررسی تکنیک هایی برای بهره برداری از تکنیک های استدلال تخصصی نظیر موارد اجراء شده در سیستم ELK هستیم تا دسته بندی هستی شناسی ها را تسریع نماییم که درون تکه OWL هستند که چنین تکنیک هایی می توانند همدل نمایند .



این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی