

بسم الله الرحمن الرحيم

پیشرفت پروژه در سه کار بوده است:

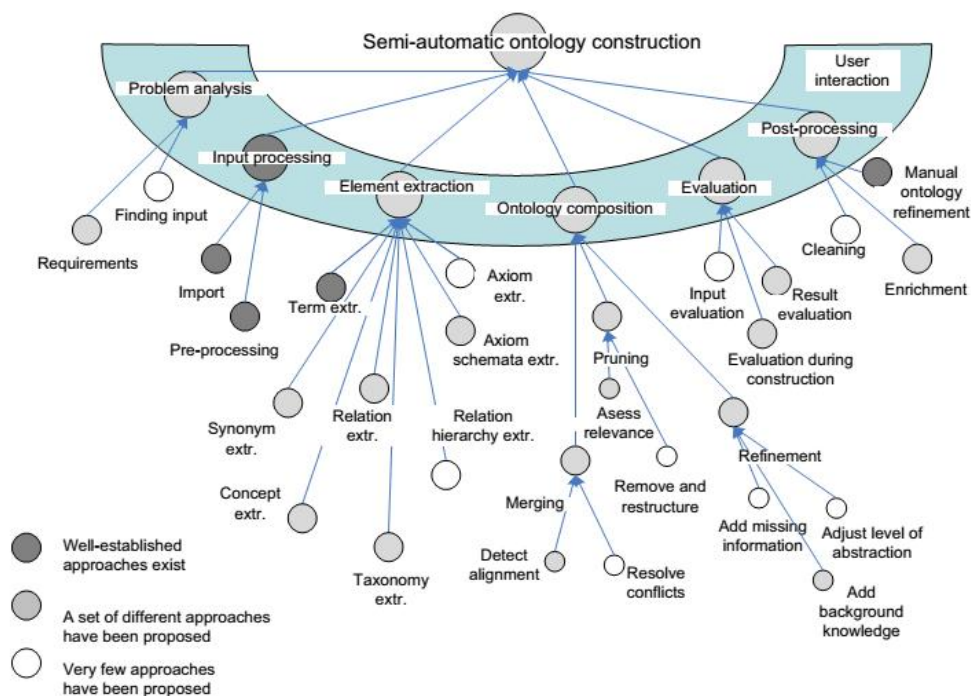
- ۱- پیشبرد پیاده سازی الگوریتم
 - ۲- تولید آنتولوژی به صورت دستی
 - ۳- مطالعه مقاله در مورد تولید خودکار آنتولوژی. توضیح مقالاتی که مطالعه شده است در ادامه آمده است.
- در مقالاتی که تا کنون دیده ام مقاله ای که به طور کاملاً خودکار و با دقت بالا به ساخت آنتولوژی بپردازد وجود نداشت. با این حال روش هایی که برای این کار به صورت نیمه خودکار ارائه شده بودند روش های جالب و قابل تاملی بودند. در مورد روشهای ماشینی کار با آنتولوژی اصطلاحی که وجود دارد یادگیری ماشین یا ontology learning است که این اصطلاح را که در مورد ontology extraction ، ontology generation و ontology acquisition به کار می برند. این موارد شامل استخراج ترم های مربوط به یک دامنه و روابط بین آنها از یک corpus از متون زبان طبیعی و سپس کد کردن آنها در قالب یک زبان آنتولوژی است. پردازش های زبان طبیعی که برای استخراج ترم ها انجام می شود شامل POS Tagging و phrase chunking است. سپس تکنیک های آماری و نمادی نیز روی آنها اعمال می شود تا ارتباط آنها استخراج شود که این کار معمولاً مبتنی بر روش های ^[4] pattern-based یا definition-based یا hypernym extraction است.

این فرآیند معمولاً در هشت گام شکسته می شود که عبارتند از:

1. [Domain terminology extraction](#)
2. [Concept discovery](#)
3. [Concept hierarchy derivation](#)
4. [Learning of non-taxonomic relations](#)
5. [Rule discovery](#)
6. [Ontology population](#)
7. [Concept hierarchy extension](#)
8. [Frame and event detection](#)

Eva Blomqvist, “Semi-automatic Ontology Construction based on Patterns”, Linköping Studies in Science and Technology, 2009

یکی از افرادی که در باره تولید خودکار آنتولوژی و در حوزه وسیع تر وب معنایی مطالعه کرده است خانم Eva Blomqvist استاد دانشگاه Linköping University در سوئد است. ایشان که در سال ۲۰۰۹ دفاع کرده است موضوع تزش ساخت نیمه خودکار آنتولوژی است. با این حال مطالعات ایشان در حال حاضر روی سیستم های تصمیم یار مبتنی بر وب معنایی و نیز linked data متمرکز شده است. تز ایشان نکات جالب و قابل تاملی دارد که در شروع کار و احتمالا در طول کار نیز بتوان از ایده های ایشان استفاده کرد. یکی از روش هایی که ایشان در تقسیم بندی کارشان و نیز برای بیان نوآوری کار خودشان و مرور متون ارائه کرده است شکلی است که این کار را به صورت آنتولوژی نشان میدهد.



در این شکل علاوه بر این که تقسیم بندی شده است روی روشهایی که قبلا کار شده است و مرور متون شده است نوآوری های خودش را نیز بیان می کند.

آنچه بیان می کند این است که بعد از استخراج مفاهیم کاری که باید انجام شود گذاشتن آنها کنار همدیگر و سپس prun کردن و تصفیه آن است. گذاشتن مفاهیم کنار همدیگر همانند کاری است که در تطبیق و مرج

کردن آنتولوژی انجام می شود و بنابراین می توان از ایده هایی که آنها وجود دارد نیز استفاده کرد. Prun کردن نیز شامل ارزیابی ارتباط نودهای آنتولوژی و نیز باز سازی آن در صورتی که بعضی بخش های آن حذف شوند است. تصفیه نیز شامل انجام کارهایی برای شناسایی این است که آیا مفهوم مهمی جا نیفتاده است و اگر افتاده اضافه کردن آن به آنتولوژی .

الگوریتم هایی که در یادگیری آنتولوژی مطرح شده است در چند لایه قابل تفکیک هستند که به لایه های کیک یادگیری آنتولوژی معروف است

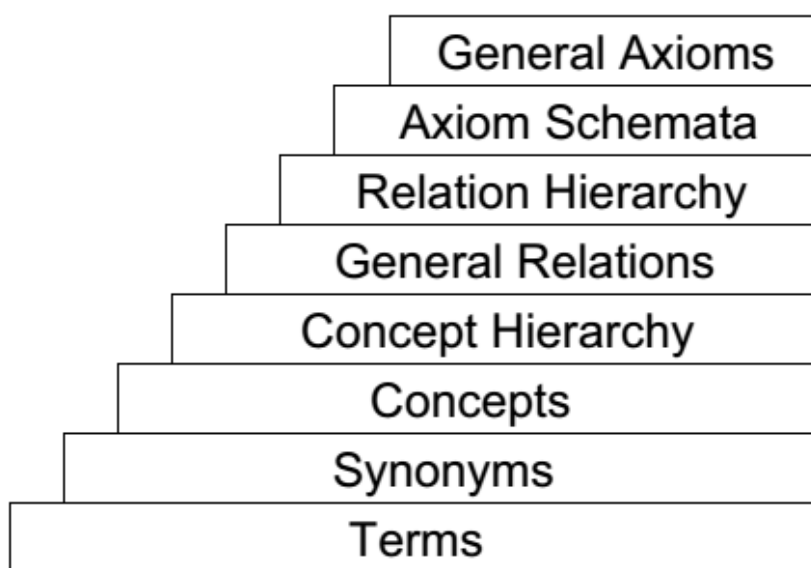


Figure 2.5: The Ontology Learning layers.

با توجه به پیشینه بیان می کند که بیشتر روش هایی که در یادگیری آنتولوژی به کار می رود مبتنی بر دانش های NLP، داده کاوی، پردازش متن و یادگیری ماشین است. کاری که خود ایشان انجام داده است مبتنی بر الگو است.

Christian Blaschke, Alfonso Valencia, “Automatic Ontology Construction from the Literature”, Genome Informatics 13: 201– 213 (2002)

ایده ای که این مقاله داشت کلاستر کردن مفاهیم است که البته روی ساختار های پروتئینی انجام شده است. روش کار بدین صورت است که مفاهیم را کلاستر کرده است و نیز از یک امتیاز هم برای آنها استفاده کرده است. یک ماتریس فاصله نیز تشکیل داده است که فاصله هر دو کلاستر را با هم نمایش می دهد. در هر مرحله دو کلاستری که کمترین فاصله را دارند را ادغام می کند و کلاستر جدید را به جای آن وارد ماتریس کرده و اطلاعات ماتریس را به روز می کند و این مرحله هر بار تکرار می شود تا کل آنتولوژی ساخته شود.