Use of Existing Ontology Elements in Ontology Construction:
Presenting and Testing a Systematic Merge-based Method

Omid Milanifard*
PhD in Information Technology Management; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran Email: omid@um.ac.ir

Mohsen Kahani
PhD in Computer Engineering; Professor of Computer Engineering; Ferdowsi University of Mashhad; Visiting Professor; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran Email: kahani@um.ac.ir

Mahdi Alipour Hafezi
PhD in Knowledge and Information Science; Assistant Professor of Information Science and Knowledge Studies Group; Allameh Tabataba'i University; Tehran, Iran; Visiting Professor; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran Email: meh.hafezi@gmail.com

Iranian Journal of Information Processing and Management

Received: 27, Feb. 2018 **Accepted: 27, May 2020**

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 36 | No. 3 | pp. 655-678
Spring 2021

Abstract: Integration of existing ontologies is an approach for reusing ontology elements in construction of a new ontology. Researchers have proposed various tools and methods for ontology merging, but very few of them work in an integrated manner with a comprehensive ontology construction methodology and also none of them support Persian language. In this paper we proposed a systematic merge-based method for reusing ontology elements with Persian support in ontology construction.

Proposed method built based on the design science research method. The process and algorithms of proposed method implemented and improved based on general cycle of design science method. Proposed method used to create a reference ontology for educational and research organizations in the subset of Iranian ministry of science, research and technology. Final product was evaluated and verified by a gold standard. An ontology merge method based on reuse of ontology elements and a higher education ontology are main products of this research.

Keywords: Ontology Development, Reusing Ontology, Ontology Merging, Ontology Construction Method

* Corresponding Author

استفاده از عناصر هستان‌نگاری موجود در ساخت هستان‌نگار جدید: ارائه و آزمون روشی نظام‌مند مبتنی بر ادغام هستان‌نگارها

امید میلانی‌فرد
دکتری مدیریت فناوری اطلاعات؛ پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران؛ پدیدآور رابط omid@um.ac.ir

محسن کاهانی
دکتری مهندسی کامپیوتر؛ استاد؛ گروه کامپیوتر؛ دانشگاه فردوسی مشهد، ایران؛ استاد مدعو پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران kahani@um.ac.ir

مهدی علیپور حافظی
دکتری علم اطلاعات و دانش‌شناسی؛ استادیار؛ گروه علم اطلاعات و دانش‌شناسی؛ دانشگاه علامه طباطبایی؛ تهران، ایران؛ استاد مهمان پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران؛ meh.hafezi@gmail.com

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۳۹۶/۱۲/۰۸ پذیرش: ۱۳۹۹/۰۳/۰۷ مقاله برای اصلاح به مدت ۵۳ روز نزد پدیدآوران بوده است.

چکیده: یکی از رویکردها در ساخت هستان‌نگار جدید ادغام عناصر هستان‌نگارهای موجود است. به‌رغم ارائهٔ ابزار و روش‌های مختلف ادغام هستان‌نگار، تعداد بسیار کمی از آن‌ها در تعاملی یکپارچه با یک روش‌شناسی توسعهٔ هستان‌نگار عمل کرده و هیچ‌یک زبان فارسی را پشتیبانی نمی‌کنند. در این نوشتار روشی نظام‌مند برای استفاده از هستان‌نگارهای موجود با پشتیبانی از زبان فارسی و در تعامل با یک روش‌شناسی کامل توسعهٔ هستان‌نگار معرفی شده است. از آنجا که این پژوهش به‌دنبال ارائهٔ یک محصول فناوری اطلاعات (یک روش و ابزار مرتبط با آن) است، روش پژوهش علم طراحی را به‌کار گرفته است. فرایند و الگوریتم‌های روش پیشنهادی این پژوهش بر اساس چرخهٔ عمومی پژوهش علم طراحی ایجاد شده و بهبود و تکامل یافته‌اند. این روش برای تولید هستان‌نگار جامع مؤسسه‌های آموزشی و پژوهشی زیرمجموعهٔ وزارت علوم، تحقیقات و فناوری مورد استفاده قرار گرفت و محصول نهایی در تقابل با یک استاندارد طلایی، ارزیابی و صحت آن تأیید شد. یافته‌های

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۳۶ | شمارهٔ ۳ | ص‌ص ۶۵۵-۶۷۸
بهار ۱۴۰۰

مهم این پژوهش شامل یک روش ساخت هستان‌نگار با رویکرد ادغام عناصر هستان‌نگارهای موجود و همچنین، یک هستان‌نگار آموزش عالی حاصل از اجرای این روش است.

کلیدواژه‌ها: توسعهٔ هستان‌نگار، استفادهٔ مجدد از هستان‌نگارها، ادغام هستان‌نگار، روش تولید هستان‌نگار

۱. مقدمه

هستان‌نگار[1]، توصیفی صریح از مفاهیم یک حوزه است که اجماع نظر خبرگان در مورد آن وجود دارد و به زبانی بیان می‌شود که توسط ماشین قابل خواندن باشد (Gruber 1995). منظور از هستان‌نگار در فناوری اطلاعات، یک قلم اطلاعاتی است که دانش یک حوزهٔ خاص را مدل می‌کند (Guarino, Oberle and Staab, 2009) و عناصر اصلی آن شامل کلاس (معادل مفهوم در عالم واقع)، روابط سلسله‌مراتبی بین کلاس‌ها، خصوصیات داده[2] (بیان‌کنندهٔ ویژگی‌های کلاس) و خصوصیات شیء[3] (روابط غیر سلسله‌مراتبی بین کلاس‌ها)ست.

ساخت هستان‌نگار در عمل کاری پرهزینه و دشوار است. یکی از راه‌های افزایش کیفیت و کاهش هزینهٔ ساخت که در برخی از پُراستنادترین روش‌های توسعهٔ هستان‌نگار مانند انترپرایز آنتولوژی[4] (Uschold and Gruninger 1996)، مت‌آنتولوژی[5] (Fernández-López and Gómez-Pérez 2002)، توسعهٔ هستان‌نگار ۱۰۱ (Noy and McGuinness 2001) و نئون[6] (Suarez-Figueroa 2010) مورد اشاره قرار گرفته، استفادهٔ مجدد از هستان‌نگارهای موجود است. به‌طور کلی، رویکرد ادغام هستان‌نگارهای موجود برای استفادهٔ مجدد از آن‌ها مورد توجه این روش‌ها بوده، اما به جزئیات چگونگی انجام کار پرداخته نشده است. به همین دلیل و با توجه به محدودیت‌های موجود در روش‌های موجود ادغام هستان‌نگارها، نیاز به روشی جدید و دارای جزئیات کافی و یکپارچه با یک روش‌شناسی توسعهٔ هستان‌نگار برای این موضوع احساس می‌شود. هدف از پژوهش جاری طراحی و آزمون روشی نظام‌مند با جزئیات کافی و پشتیبانی از زبان فارسی برای استفادهٔ مجدد از هستان‌نگارهای موجود در ساخت یک هستان‌نگار جدید است.

ویژگی‌های مورد انتظار از هدف ذکرشده عبارت‌اند از:

1. ontology
2. data property
3. object property
4. enterprise ontology
5. methontology
6. Neon

۱. روش طراحی‌شده چارچوب فعالیت‌های مورد نیاز برای تجمیع هستان‌نگارها به قصد استفادهٔ مجدد را مشخص کرده باشد؛
۲. الگوریتم انجام هر فعالیت با جزئیات کافی مشخص باشد؛
۳. روش طراحی‌شده نقاط ضعف اصلی روش‌های موجود ادغام هستان‌نگار را نداشته باشد.

از نتایج این پژوهش در شکل‌دهی به یک روش‌شناسی جامع ساخت هستان‌نگار فارسی با نام «انتیراندااک»[1] (میلانی‌فرد و کاهانی ۱۳۹۷) استفاده شده است. الگوریتم‌های معرفی‌شده در این نوشتار به‌صورت یک پیمانه[2] نرم‌افزاری در سامانهٔ یکپارچهٔ توسعهٔ مشارکتی هستان‌نگار «انتیراندااک»، پیاده‌سازی و آزمون شده‌اند.

در ادامه، ابتدا پیشینهٔ پژوهش مرور شده و سپس، روش مورد استفاده معرفی می‌شود. پس از این دو بخش، روش پیشنهادی با ذکر جزئیات الگوریتم‌های آن تشریح شده است. در انتهای این نگارش، پیاده‌سازی عملی و نتایج ارزیابی محصول نهایی حاصل از اجرای روش پیشنهادی بیان شده است.

۲. پیشینهٔ پژوهش

برای مسئلهٔ ادغام هستان‌نگارها پژوهش‌های زیادی صورت گرفته و روش‌های مختلفی پیشنهاد شده است. از جمله نگارش‌های پراستنادی که فهرست جامعی از این روش‌ها ارائه کرده‌اند، می‌توان به (Shvaiko and Euzenat 2013) و (Choi, Song and Han 2006) اشاره کرد. رویکردهای مورد استفاده در روش‌های معرفی‌شده را می‌توان به شکل زیر دسته‌بندی کرد (Ramar and Gurunathan 2016):

۱. مبتنی بر ساختار[3]: در این رویکرد از ساختار هستان‌نگارهای ورودی مانند روابط سلسله‌مراتبی، مفاهیم همنیا[4] و ارتباطات بین گره‌ها در گراف هستان‌نگار (روابط غیر سلسله‌مراتبی) استفاده می‌شود. در این رویکرد فرض می‌شود که هستان‌نگارها گراف‌های برچسب‌دار هستند و مشابهت دو مفهوم بر اساس تحلیل جایگاه آن دو مفهوم در گراف پیشنهاد می‌شود. از روش‌های مشهوری که از این رویکرد استفاده

1. ontirandoc 2. module 3. structure based 4. sibling concepts

کرده‌انـد می‌تـوان بـه «انکورپرامپـت»[1] (Noy and Musen 2001) اشـاره کـرد؛

۲. مبتنـی بـر مشـابهت شـکلی[2]: در ایـن رویکـرد از مقایسـهٔ رشـته‌های حرفـی اسـتفاده می‌شـود. ایـن رویکـرد کـه می‌توانـد بـر روی نـام و یـا برچسـب‌های عناصـر هسـتان‌نگار اعمـال شـود، بـر پایـهٔ ایـن فـرض اسـت کـه دو عنصـر هسـتان‌نگاری کـه نـام یـا برچسـب آن‌هـا از نظـر شـکلی مشـابهت دارد، احتمـالاً مشـابه هسـتند. از جملـه روش‌هـای پراسـتنادی کـه از ایـن رویکـرد اسـتفاده کرده‌انـد، می‌تـوان بـه «کُمـا»[3] (Do and Rahm 2002) اشـاره کـرد؛

۳. مبتنـی بـر نمونـه[4]: در ایـن رویکـرد فـرض می‌شـود چنانچـه نمونه‌هـای دو مفهـوم دارای همپوشـانی قابـل توجهـی باشـند، احتمـال مشـابهت آن‌هـا وجـود دارد. از جملـه روش‌هایـی کـه از ایـن رویکـرد اسـتفاده کرده‌انـد، می‌تـوان بـه «گلـو»[5] (Doan et al. 2004) اشـاره کـرد؛

۴. مبتنـی بـر دانـش پیش‌زمینـه[6]: در اکثـر روش‌هـا و ابـزار، تنهـا بـه دانـش نهفتـهٔ داخـل خـود هسـتان‌نگار (کلاس‌هـا، روابـط، خصوصیـات، قیـود[7] و نمونه‌هـا) اکتفـا شـده اسـت. انـدک روش‌هایـی کـه بـه اسـتفاده از دانـش پیش‌زمینـه در بهبـود سـاخت محصـول نهایـی توجـه داشـته‌اند، محـدود بـه اسـتفاده از دانـش موجـود در هسـتان‌نگارهای سـطح بـالا[8] ماننـد (Menéndez-Mora & Ichise 2013)، دانـش نهفتـه در پیکـره[9] ماننـد (Madhavan et al. 2005) و وب معنایـی[10] ماننـد (Jain et al. 2010) بوده‌انـد.

در جـدول ۱، پژوهش‌هـای صورت‌گرفتـه در زمینـهٔ ادغـام هسـتان‌نگارها (در بـازهٔ زمانـی سـال ۲۰۱۰ بـه بعـد) کـه منجـر بـه ارائـهٔ روش شـده‌اند، فهرسـت شـده اسـت.

---

1. AnchorPrompt
2. terminological based
3. COMA
4. instance
5. GLUE
6. background knowledge based
7. restrictions

۸. هسـتان‌نگارهای سـطح بـالا (upper ontology) هسـتان‌نگارهایی هسـتند کـه واژگان بسـیار عمومـی درون تمـام حوزه‌هـا (ماننـد شـی، خصوصیـت، رابطـه) را مـدل می‌کننـد.

9. corpus
10. Semantic Web

جدول ۱. پژوهش‌های پیشین که منجر به ارائهٔ روش برای ادغام هستان‌نگار شده‌اند

نام روش	استناد	مبتنی بر مشابهت شکلی	مبتنی بر ساختار	مبتنی بر نمونه
ری‌مام[1]	(Li et al. 2008)	X	X	X
کدی[2]	(Huber et al. 2011)	X	X	X
سرومپ[3]	(Ba and Diallo 2013)	X	X	
مس‌اس‌اس[4]	(Cheatham and Hitzler 2013)	X	X	
لاگ‌مپ[5]	(Jiménez-Ruiz et al. 2014)	X	X	
یام++[6]	(Ngo and Bellahsene 2012)	X	X	X
اروما[7]	(David 2011)	X		X

همان‌طور که در جدول ۱، مشاهده می‌شود، اکثر پژوهش‌ها رویکردهای مبتنی بر مشابهت شکلی و ساختار را مورد استفاده قرار داده‌اند. مطالعهٔ پژوهش‌های پیشین نشانگر وجود محدودیت‌هایی به‌شرح زیر است:

۱. اکثر روش‌های ارائه‌شده به‌عنوان بخشی از یک فعالیت پژوهشی و متناسب با نیازهای آن پژوهش طراحی شده است (Otero-Cerdeira, Rodríguez-Martínez and Gómez-Rodríguez 2015). به همین دلیل، پس از مدتی منسوخ شده و دیگر کاری برای بهبود و رفع اشکالات آن‌ها انجام نشده است. به‌عنوان مثال، ابزار پرامپت[8] که بین سال‌های ۲۰۰۰ و ۲۰۰۵ پیشگام ابزار ادغام هستان‌نگار بود، با وجود داشتن اشکالات از حدود ۱۰ سال پیش دیگر به‌روز نشده و در نسخه‌های جدید ویرایشگر هستان‌نگار پروتژه[9] از آن پشتیبانی نمی‌شود؛

۲. از محدودیت‌های ابزار موجود عدم پشتیبانی آن‌ها از زبان فارسی در مشابهت‌یابی معنایی توسط شبکهٔ واژگان است. به‌عبارت دیگر، هیچ‌یک از ابزار یادشده از شبکهٔ واژگان فارسی استفاده نکرده است؛

۳. در بیشتر روش‌ها و ابزار معرفی‌شده از دانش پیش‌زمینه برای بهبود فرایند ساخت استفاده نشده است.

1. RiMOM
2. CODI
3. SerOMap
4. MassSS
5. LogMap
6. YAM++
7. AROMA
8. PROMPT
9. Protégé: http://protege.stanford.edu.

۳. روش پژوهش

با توجه به این‌که موضوع این پژوهش یک موضوع کاربردی فناوری اطلاعات است، روش علم طراحی[1] برای انجام آن مورد استفاده قرار گرفت. پژوهش علم طراحی که به‌عنوان پژوهش سازنده[2] نیز خوانده می‌شود، نوعی از تولید دانش علمی است که همزمان با مشارکت در تولید علم، به‌دنبال توسعهٔ سازه‌های خلاقانه به‌منظور حل مسائل دنیای واقعی است (Dresch, Lacerda and Antunes 2015). این روش در دستهٔ روش‌های کیفی در حوزهٔ مطالعات بین رشته‌ای قرار می‌گیرد و پژوهش‌های صورت‌گرفته با آن عموماً بدون فرضیه‌ای برای اثبات بوده و بیشتر به‌دنبال راه‌حل‌های عملی برای مسائل دنیای واقعی هستند (Vaishnavi & Kuechler 2015).

در روش علم طراحی یک چرخه برای حل مسئلهٔ پژوهشی به‌صورت مداوم و تا رسیدن به جواب قابل قبول تکرار می‌شود. همان‌گونه که در شکل ۱، مشاهده می‌شود، این چرخه شامل آگاهی از مسئله و شناخت کامل آن، ارائهٔ پیشنهاد برای حل مسئله، اجرای راه‌حل و ارزیابی نتایج حاصل از اجرای راه‌حل است. بر اساس نتایج به‌دست‌آمده از ارزیابی اجرای راه‌حل، چالش‌ها و اشکالات مجدد بررسی می‌شود و راه‌حل مسئله اصلاح می‌گردد و مجدد مورد آزمون قرار می‌گیرد (Vaishnavi & Kuechler 2015).

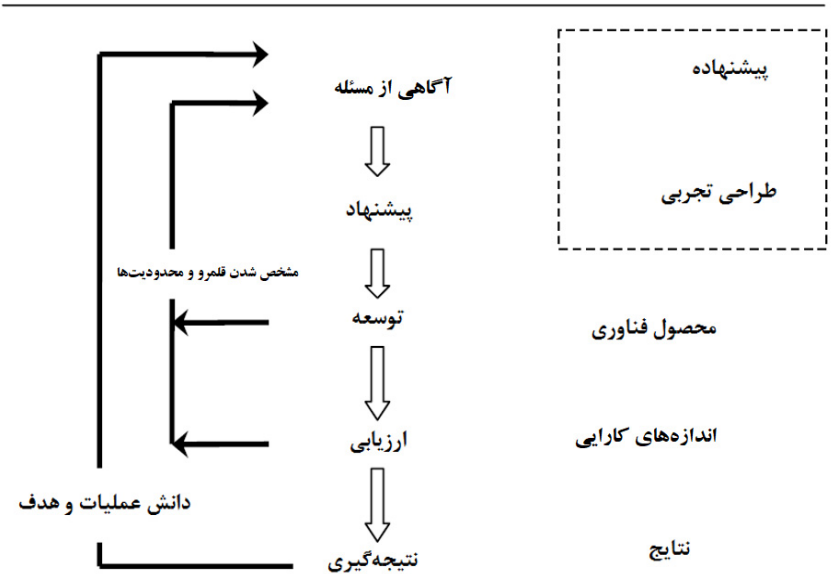
جامعهٔ مورد پژوهش، مستندات منتشرشده و روش‌های مختلف ارائه‌شده برای ادغام هستان‌نگارهاست. از این روش‌ها به‌عنوان دانش نخستین برای آگاهی از مسئله (مرحلهٔ اول در چرخه) و همچنین، کمک به تولید ایده جهت ساخت پیشنهادهای اولیه (مرحلهٔ دوم چرخه) استفاده می‌شود.

ابتدا، بر اساس مرور پیشینهٔ پژوهش، طرح اولیه‌ای برای محصول مورد نظر این پژوهش تهیه شد که به اختصار «راما»[3] نامیده می‌شود. سپس، بر اساس تکرار چرخهٔ پژوهش علم طراحی، به‌تدریج با ارائهٔ راه‌حل برای انجام هر یک از گام‌های محصول، الگوریتم‌ها و روش انجام کار بهینه‌سازی شد. برای سنجش کارایی راه‌حل‌های ارائه‌شده از روش پیاده‌سازی عملی الگوریتم‌ها و بررسی خروجی اجرای آن‌ها استفاده شد. چرخهٔ ارائهٔ راه‌حل پیشنهادی برای چالش‌های موجود در ادغام هستان‌نگارها و سپس، پیاده‌سازی

1. design science research
2. constructive research
۳. روش استفادهٔ مجدد از آنتولوژی‌های موجود

و آزمون عملی آن‌ها چندین بار تکرار شد تا در نهایت، محصول این پژوهش یعنی «راما» ایجاد گردید.

برای آزمون صحت عملکرد «راما» این روش به‌صورت عملی برای ساخت هستان‌نگار جامع مؤسسات آموزشی و پژوهشی زیرمجموعهٔ وزارت عتف[1] مورد استفاده قرار گرفت. ساخت این هستان‌نگار به سفارش دفتر فناوری اطلاعات وزارت عتف صورت گرفته و برای انجام این کار یک روش جدید با نام «انتیرانداک» ایجاد شده است (میلانی‌فرد و کاهانی ۱۳۹۷). «انتیرانداک» از سه‌سویه‌سازی[2] منابع دانشی (هستان‌نگارهای موجود، مستندات متنی و پایگاه دادهٔ عملیاتی) و روش‌های استخراج هستان‌نگار (ادغام هستان‌نگارهای موجود، استخراج هستان‌نگار از متن، مهندسی معکوس پایگاه داده) به‌منظور افزایش قابلیت اطمینان محصول خروجی استفاده می‌کند. پژوهش جاری به‌منظور ساخت روشی برای بخش اول، یعنی استفاده از منبع هستان‌نگارهای موجود و روش ادغام مناسب آن‌ها انجام شده است. لازم به توضیح است که «راما» قابلیت استفاده به‌صورت مجزا و یا به‌عنوان بخشی از روش کامل «انتیرانداک» را دارد.



شکل ۱. روش عمومی علم طراحی (Vaishnavi & Kuechler 2015)

۱. وزارت علوم، تحقیقات و فناوری

2. triangulation

٤. یافته‌ها

جریان کاری «راما»، بر اساس مطالعهٔ روش‌های موجود ادغام هستان‌نگارها و با افزودن گام برچسب‌گذاری عناصر هستان‌نگارها به‌منظور ایجاد سازگاری معنایی و به‌ویژه استفاده از لایهٔ معنایی زبان فارسی ایجاد شده است. این جریان کاری که در شکل ۲، نشان داده شده، شامل گام‌های زیر است:

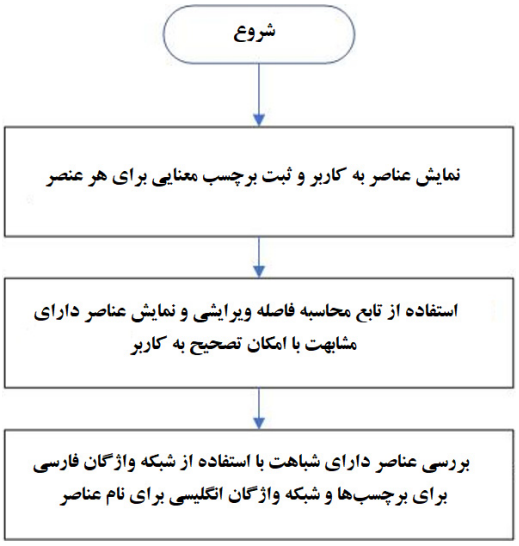
۱. شناسایی هستان‌نگارهای مرتبط با حوزهٔ مورد نظر: با بررسی روش‌های ارائه‌شده برای این کار، راهنمای گمانه‌زنی هستان‌نگار[1] که توسط کنسرسیوم جهانی وب ارائه‌شده (W3C 2010)، به‌دلیل اعتبار نهاد پیشنهاددهندهٔ آن، به‌عنوان روش انجام این گام انتخاب شده است؛

۲. برچسب گذاری عناصر هستان‌نگار: از آنجا که واژگان به‌کاررفته در نامگذاری عناصر هستان‌نگارها بین هستان‌نگارهای مختلف یکسان نیست و در برخی موارد، عنوان عناصر هستان‌نگار به‌صورت کدگذاری شده است، با برچسب‌گذاری هر یک از عناصر یک لایهٔ جدید معنایی بر روی کل عناصر هستان‌نگارها کشیده شده و واژگان به‌کاررفته برای مفاهیم یکسان سازگار می‌شود. در این روش، برچسب‌گذاری به یک زبان مشترک[2] انجام می‌شود. جریان کاری این مرحله در شکل شماره ۳، نشان داده شده است.

1. ontology dowsing

۲. در ابزار پیاده‌سازی‌شده، زبان فارسی انتخاب شده است.

شکل ۲. جریان کاری «راما»



شکل ۳. جریان کاری مرحلهٔ برچسب‌گذاری

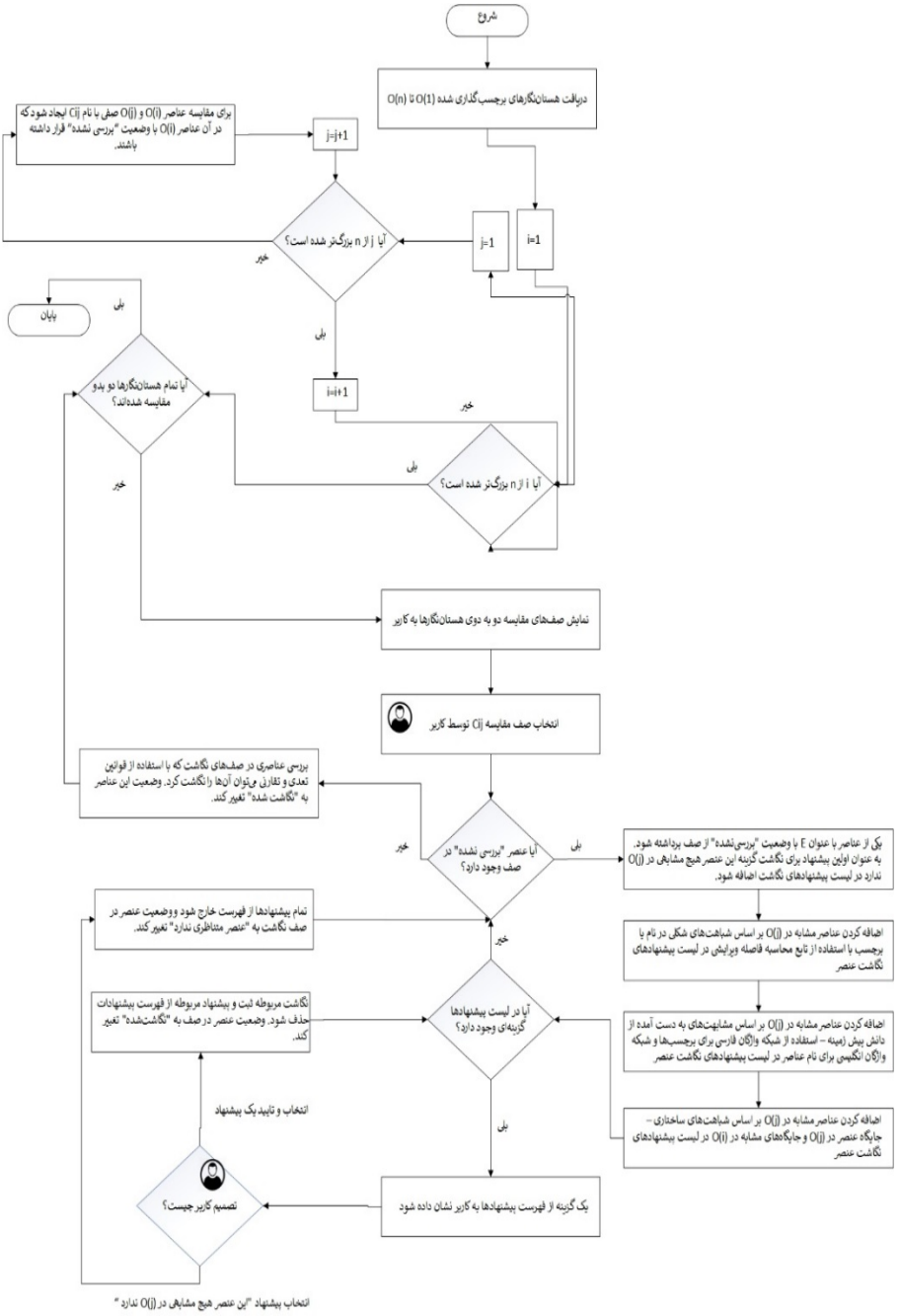
۳. نگاشت عناصر هستان‌نگاری: در این گام عناصر هستان‌نگارها به یکدیگر نگاشت می‌شوند. شکل ۴، جریان کاری مربوط به این گام را نشان می‌دهد.

به‌دلیل نامگذاری‌های متفاوت و گاهی اوقات مبهم عناصر هستان‌نگار و همچنین، تفاوت دیدگاه مدل‌سازان که منجر به مدل‌کردن یک واقعیت یکسان به چندین شکل می‌شود، یافتن عناصر مترادف در هستان‌نگارها را نمی‌توان به‌صورت کاملاً خودکار انجام داد و انجام کارآمد این گام نیازمند مداخلهٔ کاربر است (Massmann et al. 2011). فعالیت‌هایی که نیازمند مداخلهٔ کاربر است، با نشانه 👤 مشخص شده‌اند.

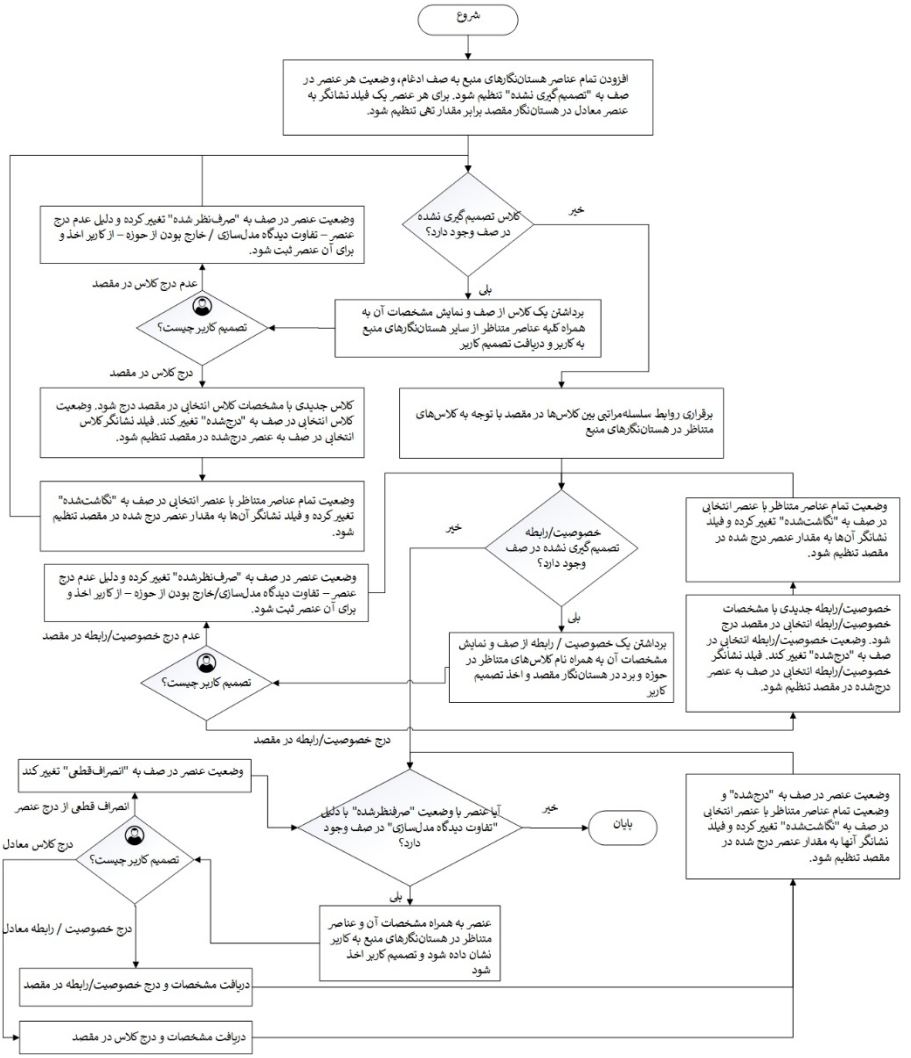
نکات زیر در این گام مورد توجه است:

- به‌دلیل دیدگاه متفاوت مدل‌سازان، یک عنصر در یک هستان‌نگار الزاماً به عنصری از همان نوع در هستان‌نگار دیگر نگاشت نمی‌شود. به بیان دیگر، ممکن است یک کلاس در هستان‌نگار اول به یک خصوصیت/ رابطه یا یک مقدار مجاز خصوصیت در هستان‌نگار دوم نگاشت شود. در فهرست عناصر همسان پیشنهادی به کاربر ابتدا باید عناصر مشابه از همان نوع و سپس، عناصر مشابه انواع دیگر به کاربر ارائه شود؛
- از آنجا که نگاشت دو عنصر به معنای رابطهٔ معادل بودن آن دو عنصر است، این رابطه خاصیت تعدّی[1] و تقارنی دارد. از این خاصیت می‌توان در تکمیل سریع‌تر ماتریس نگاشت هستان‌نگارها به یکدیگر استفاده کرد؛
- ممکن است هستان‌نگارهای مختلف یک حوزه را با دانه‌بندی‌های متفاوت مدل کرده باشند. به‌عبارت دیگر، سطح جزئیات هستان‌نگارها با هم متفاوت بوده و یک عنصر در یک هستان‌نگار ممکن است تجمیعی از چند عنصر در هستان‌نگار دیگر باشد. به همین دلیل، باید امکان نگاشت یک عنصر از یک هستان‌نگار را به بیش از یک عنصر در هستان‌نگار دیگر فراهم کرد؛
- بررسی هستان‌نگارها از نظر میزان پوشش عناصر یکدیگر: با محاسبهٔ تعداد عناصر نگاشت‌شده به کل عناصر مشخص می‌شود که هر هستان‌نگار با چه درصدی کلاس‌ها و خصوصیات سایر هستان‌نگارها را پوشش می‌دهد. چنانچه میزان پوشش یک هستان‌نگار با سایر هستان‌نگارها از مقدار آستانه‌ای کمتر باشد، می‌تواند نشان‌دهندهٔ میزان کم ارتباط آن با حوزهٔ موضوعی هستان‌نگار هدف باشد و به حذف آن از مجموعهٔ هستان‌نگارهای مورد استفاده منجر شود.

1. transitive



شکل ٤. جریان کاری گام نگاشت عناصر هستان‌نگارها



شکل ۵. الگوریتم ادغام هستان‌نگارها

۴. ادغام هستان‌نگارها: الگوریتم طراحی‌شده در این گام که در شکل ۵، نشان داده شده، مانند گام نگاشت عناصر هستان‌نگاری، نیازمند مداخلهٔ کاربر بوده و به‌شرح زیر است:

◇ ابتدا، مخزنی از تمام عناصر هستان‌نگارهای منبع به‌صورت یک صف تشکیل می‌شود. یکی از موضوعات مهم در قابلیت به‌کارگیری یک هستان‌نگار داشتن مستندات کافی در خصوص عناصر آن است (Zemmouchi-Ghomari & Ghomari

2013). به همین دلیل، برای هر عنصر در صف یک نشانگر به عنصر معادل در هستان‌نگار مقصد در نظر گرفته می‌شود. با داشتن چنین نشانگری منبع ایجاد هر یک از عناصر هستان‌نگار مقصد مشخص است و استفاده‌کنندهٔ نهایی می‌تواند در صورت نیاز برای درک بهتر آن عنصر به سرچشمهٔ ورود آن عنصر و فهرست تمام عناصر هستان‌نگاری متناظر آن در هستان‌نگارهای منبع، دسترسی پیدا کند؛

- از آنجا که کلاس‌ها مدلی از مفاهیم در دنیای واقعی هستند و بدون وجود آن‌ها خصوصیات و روابط، معنایی نخواهند داشت، ابتدا باید به وضعیت آن‌ها در صف رسیدگی شود. کلاس‌های تصمیم‌گیری‌نشده به همراه فهرست عناصر متناظر در سایر هستان‌نگارهای منبع به کاربر نمایش داده می‌شود. مشاهدهٔ عناصر متناظر به کاربر کمک می‌کند تا از دیدگاه سایر مدل‌سازان در خصوص آن عنصر آگاه شود. البته، باید توجه داشت که این فهرست یکی از منابع تصمیم‌گیری کاربر است و تصمیم او به موارد دیگری همچون تجربهٔ شخصی یا کاربرد مورد انتظار از آن مفهوم نیز وابسته است. تصمیم کاربر در خصوص هر کلاس می‌تواند یکی از دو مورد زیر باشد:

الف) درج کلاس در مقصد: در این صورت وضعیت عنصر در صف به «درج‌شده» تغییر کرده و کلاس معادلی در هستان‌نگار مقصد با عنوان و برچسب عنصر منبع درج می‌شود. نشانگر عنصر در صف نیز به کلاس درج‌شده تنظیم می‌شود. زمانی که یک کلاس معادل در هستان‌نگار مقصد درج می‌شود، در واقع، آن مفهوم در هستان‌نگار مقصد در نظر گرفته شده و دیگر نیازی به ورود مفاهیم یا خصوصیات متناظر به هستان‌نگار مقصد نیست. به همین دلیل، تمام عناصر متناظر در صف به‌صورت خودکار به وضعیت «نگاشت‌شد» تغییر می‌کنند و نشانگر آن‌ها به کلاس جدید تنظیم می‌شود؛

ب) عدم درج کلاس در مقصد: کاربر به دو دلیل امکان دارد از درج کلاس معادلی در هستان‌نگار مقصد خودداری کند:

۱. خارج از موضوع بودن آن کلاس در قلمرو تعریف‌شده برای هستان‌نگار نهایی: هستان‌نگار، تخمینی از یک حوزهٔ دانشی یا کسب‌وکار است و در ارزیابی آن باید درجهٔ چنین تخمینی اندازه‌گیری شود. به همین منظور،

دو سنجهٔ دقت[1] و جامعیت[2] پیشنهاد شده است (Guarino 2004). منظور از دقت این است که هستان‌نگار عناصر غیر مرتبط با موضوع را شامل نشود و جامعیت هم به این معناست که هستان‌نگار تمام مفاهیم حوزه را دربرگیرد. رد کردن کلاس‌های غیر مرتبط در این مرحله سنجهٔ دقت هستان‌نگار مقصد را بهبود خواهد داد؛

۲. تفاوت دیدگاه مدل‌سازی: همان‌طور که پیش‌تر گفته شد، ممکن است کاربر به‌دلیل کاربرد هستان‌نگار نهایی یا تجربهٔ شخصی بخواهد مفهوم ذکرشده در منبع را به شکل دیگری مانند خصوصیت یا رابطه در هستان‌نگار مقصد مدل کند.

◇ پس از تصمیم‌گیری در خصوص کلاس‌های موجود در صف، بر اساس روابط سلسله‌مراتبی موجود در هستان‌نگارهای منبع، روابط سلسله‌مراتبی در هستان‌نگار مقصد محاسبه و درج می‌شود. به این منظور، تمام روابط سلسله‌مراتبی در هستان‌نگارهای منبع بررسی می‌شوند و در صورتی که کلاس پدر و کلاس فرزند هر دو دارای متناظری در هستان‌نگار مقصد باشند، بین دو کلاس در هستان‌نگار مقصد رابطهٔ سلسله‌مراتبی برقرار می‌شود؛

◇ پس از ایجاد نسخهٔ اولیه از هستان‌نگار مقصد، که تنها شامل کلاس‌ها و روابط سلسله‌مراتبی بین آن‌هاست، نوبت بررسی خصوصیات/ رابطه‌های موجود در صف منبع می‌رسد. در این مرحله نیز به‌ترتیب، یک خصوصیت/ رابطه از صف برداشته شده و با نمایش عناصر متناظر آن در هستان‌نگارهای منبع از کاربر خواسته می‌شود در خصوص درج یا عدم درج آن در هستان‌نگار مقصد تصمیم‌گیری کند. در زمان نمایش خصوصیت/ رابطه باید کلاس‌های متناظر با کلاس‌های دامنه[3] و بُرد[4] به کاربر نمایش داده شود؛

◇ پس از تعیین وضعیت تمام عناصر صف، عناصری که به‌دلیل تفاوت دیدگاهِ مدل‌سازی به شکل موجود در منبع به مقصد اضافه نشده‌اند، مجدد بررسی شدند. با نمایش هر یک از این عناصر به همراه موارد متناظر آن از کاربر خواسته می‌شود در خصوص درج عنصر معادلی به شکل دلخواه در هستان‌نگار مقصد و یا صرف نظر قطعی از آن عنصر تصمیم‌گیری کند.

1. precision 2. recall 3. domain 4. range

۵. پالایش هستان‌نگار مقصد: هستان‌نگار اولیهٔ حاصل از ادغام به‌دلیل تفاوت دانه‌بندی و مدل‌سازی هستان‌نگارهای منبع دارای اشکالاتی خواهد بود که عملیات زیر برای اصلاح آن پیشنهاد می‌شود:

◇ شناسایی و بازبینی روابط مشابه: چنانچه بین دو کلاس بیش از یک رابطه وجود داشته باشد، این روابط به‌صورت گروه‌بندی‌شده به کاربر نمایش داده شوند تا کاربر در صورت نیاز، روابط را با هم ادغام و یا روابط اضافی را حذف کند؛

◇ شناسایی خصوصیاتی که برای دو کلاسِ دارای رابطهٔ سلسله‌مراتبی تکرار شده‌اند: در زمان ادغام هستان‌نگارها امکان دارد در دامنهٔ یک خصوصیت، کلاس پدر و فرزند با هم حضور داشته باشند. از آنجا که کلاس‌های فرزند به‌صورت پیش‌فرض، خصوصیات کلاس پدر را به ارث می‌برند، این موارد باید به کاربر نمایش داده شوند تا اصلاح لازم را انجام دهد؛

◇ شناسایی روابطی که برای دو کلاس در یک رابطهٔ سلسله‌مراتبی تکرار شده‌اند: آن دسته از خصوصیات شیء در هستان‌نگار مقصد که در دامنه یا بُرد آن‌ها دو کلاس دارای رابطهٔ سلسله‌مراتبی وجود داشته باشد، برای تصمیم‌گیری و اصلاح لازم به کاربر نشان داده شوند؛

◇ پیشنهاد روابط سلسله‌مراتبی: مهم‌ترین رابطهٔ معنایی که بین مفاهیم در یک هستان‌نگار مدل می‌شود، رابطهٔ سلسله‌مراتبی[1] است. چنانچه این رابطه بین مفهوم اول (پدر) و مفهوم دوم (فرزند) برقرار باشد، مشخص می‌شود که مفهوم دوم نوعی از مفهوم اول است که خصوصیات و روابط مفهوم اول را به ارث برده است. در این گام کلاس‌هایی که درصد خصوصیات/ روابط مشترکی بالاتر از یک مقدار آستانه[2] دارند، در دسته‌های مجزا به کاربر نمایش داده می‌شوند. کاربر می‌تواند یکی از تصمیمات زیر را اجرا کند:

الف) اگر از نظر مفهومی بتوان یکی از کلاس‌ها را به‌عنوان کلاس بالاتر در نظر گرفت، خصوصیات و روابط مشترک را از سایر کلاس‌ها حذف و رابطهٔ سلسله‌مراتبی بین آن‌ها و کلاس بالاتر برقرار کند؛

1. Is-a relation
2. threshold

ب) اگر همهٔ کلاس‌ها از نظر مفهومی در یک سطح بودند، تجمیع خصوصیات و روابط مشترک را می‌توان به‌عنوان یک کلاس جدید در نظر گرفت و با حذف خصوصیات/ روابط مشترک از همهٔ کلاس‌های دسته، یک رابطهٔ سلسله‌مراتبی بین آن‌ها و کلاس جدید ایجاد کرد؛

ج) ممکن است تجمیع خصوصیات/ روابط مشترک یک کلاس دارای مفهوم نشود که در این حال، نیازی به ایجاد رابطهٔ سلسله‌مراتبی وجود نخواهد داشت.

برای آزمون «راما»، الگوریتم‌های آن پیاده‌سازی شده و به‌صورت عملی برای ساخت هستان‌نگار جامع مؤسسه‌های آموزشی و پژوهشی وزارت عتف مورد استفاده قرار گرفت. بر اساس گام اول، با استفاده از راهنمای گمانه‌زنی هستان‌نگار، هستان‌نگارهای دارای مشابهت با حوزهٔ هدف شناسایی شدند (جدول ۲).

جدول ۲. هستان‌نگارهای موجود در حوزهٔ مؤسسات آموزشی و پژوهشی

نام هستان‌نگار	تعداد کلاس‌ها	تعداد خصوصیات/ روابط
هستان‌نگار مرجع آموزش عالی (هرو)[1]	۵۶	۱۰۲
هستان‌نگار معیار دانشگاه مهای (الیوام‌بی)[2]	۴۰	۲۶
وب معنایی برای انجمن‌های پژوهشی (اس‌دبلیوآرسی)[3]	۵۳	۴۷
هستان‌نگار دانشگاه تورنتو[4]	۵۱	۳۴
هستان‌نگار دانشگاه[5]	۶۹	۴۴
هستان‌نگار حوزهٔ آموزشی و پژوهشی (وی‌آی‌وی‌اُ)[6]	۳۸۳	۱۶۳
چارچوب مشترک اطلاعات تحقیقاتی اروپا (سریف)[7]	۲۰۷	۱۸۳

در گام دوم، تمامی عناصر هستان‌نگارها با استفاده از الگوریتم بیان‌شده در شکل ۳، برچسب‌گذاری فارسی شده، و در گام سوم، تناظریابی عناصر هستان‌نگارها

1. HERO (Higher Education Reference Ontology): http://sourceforge.net/projects/heronto/
2. LUMB (Lehigh University Benchmark): http://swat.cse.lehigh.edu/onto/univ-bench.owl
3. SWRC (Sematic Web for Research Communities): http://swrc.ontoware.org/ontology
4. http://www.cs.toronto.edu/semanticweb/maponto/MapontoExamples/univ-cs.owl
5. http://www.webkursi.lv/luweb05fall/resources/university.owl
6. VIVO: http://vivoweb.org/files/vivo-isf-public-1.6.owl
7. CERIF: http://www.dfki.de/~brigitte/CERIF/CERIF2004_1.1FDM/cerif.owl

نسبت به یکدیگر انجام شد که نتیجه در جدول ۳، آمده است[1]. در این جدول، عدد اول هر سلول، درصد کلاس‌هایی از هستان‌نگار ستون که دارای متناظری در هستان‌نگار ردیف هستند و عدد دوم، درصد پوشش خصوصیات/ روابط از هستان‌نگار ستون توسط هستان‌نگار ردیف را نشان می‌دهد.

با اجرای الگوریتم گام پنجم یک هستان‌نگار تجمیعی از هستان‌نگار موجود ساخته شد که شامل ۱۳۲ کلاس و ۲۳۷ خصوصیت/ رابطه است. در کنار هستان‌نگار حاصل از ادغام هستان‌نگارهای موجود دو هستان‌نگار دیگر نیز با استفاده از مستندات متنی (کمیتهٔ آماربخشی ۱۳۸۷) (شامل ۱۸۷ کلاس و ۳۱۱ خصوصیت/ رابطه) و مهندسی معکوس پایگاه داده یک سامانهٔ اطلاعاتی آموزش عالی با نام «سدف»[2] (شامل ۱۵۶ کلاس و ۵۷۹ خصوصیت/ رابطه) ساخته شدند.

جدول ۳. ماتریس تناظر هستان‌نگارها

	سریف	هرو	ال‌یوام‌بی	اس‌دبلیوآرسی	هستان‌نگار تورنتو	هستان‌نگار دانشگاه	وی‌آی‌وی‌اُ
سریف		۳۳,۹۳ ۲۲,۵۵	۵۷,۵ ۲۶,۹۲	۵۴,۷۲ ۳۲,۴۷	۵۶,۸۶ ۵۲,۹۴	۲۸,۹۹ ۴۵,۴۵	۲۱,۶۷ ۳۵,۵۸
هرو	۳,۸۶ ۱۸,۰۳		۷۰ ۳۸,۴۶	۴۵,۲۸ ۴۲,۵۵	۵۶,۸۶ ۵۲,۹۴	۳۰,۴۳ ۵۴,۵۵	۱۰,۹۷ ۳۲,۲۵
ال‌یوام‌بی	۰,۹۷ ۳,۲۸	۳۹,۲۹ ۱۰,۷۸		۵۰,۹۴ ۲۷,۶۶	۷۶,۴۷ ۷۰,۵۹	۲۱,۷۴ ۱۸,۱۸	۸,۰۹ ۹,۲
اس‌دبلیوآرسی	۱۰,۱۴ ۱۵,۳	۳۳,۹۳ ۲۰,۵۹	۶۲,۵ ۵۷,۶۹		۵۶,۸۶ ۵۸,۸۲	۱۷,۳۹ ۵۰	۱۱,۴۹ ۳۰,۶۷
هستان‌نگار تورنتو	۴,۸۳ ۱۲,۰۲	۴۱,۰۷ ۱۷,۶۵	۹۷,۵ ۸۰,۷۷	۵۶,۶ ۳۸,۳		۲۴,۶۴ ۴۳,۱۸	۹,۶۶ ۲۸,۲۲
هستان‌نگار دانشگاه	۲,۹ ۱۶,۳۹	۲۶,۷۹ ۱۶,۶۷	۳۵ ۲۶,۹۲	۲۴,۵۳ ۲۷,۶۶	۳۵,۲۹ ۲۹,۴۱		۷,۸۳ ۲۶,۹۹
وی‌آی‌وی‌اُ	۱۶,۹۱ ۲۲,۴	۳۹,۲۹ ۲۷,۴۵	۵۰ ۴۲,۳۱	۵۴,۷۲ ۴۸,۹۴	۵۰,۹۸ ۵۲,۹۴	۲۷,۵۴ ۵۴,۵۵	

در نهایت، هر سه این هستان‌نگارها با روش «راما» با هم ادغام و هستان‌نگار نهایی

---

۱. هستان‌نگارهای دیگری هم در گام اول شناسایی شدند که بعداً در گام چهارم و به‌دلیل نداشتن میزان پوشش کافی با سایر هستان‌نگارها حذف شدند. به جهت تلخیص مطلب، اطلاعات این هستان‌نگارها در جدول‌ها نیامده است.

۲. سامانه‌های دانشگاه فردوسی

با توجه به ابزارهای معرفی شده در گام پنجم پالایش شد. در زیر، برخی از نتایج حاصل از گام پالایش به‌صورت خلاصه بیان شده است:

۱. شناسایی و بازبینی روابط مشابه: با ادغام کلاس‌ها و روابط ممکن است بین دو کلاس چندین رابطه برقرار باشد که بخشی از آن‌ها از نظر مفهومی یکی باشند و باید ادغام شوند (به‌عنوان مثال، پس از ادغام، بین دو کلاس دانشجو و پایان‌نامه سه رابطه «ارائه‌دادن»، «تدوین‌کردن» و «دفاع‌کردن» وجود داشت که از نظر مفهومی می‌توان این سه رابطه را با هم یکسان در نظر گرفت)؛

۲. شناسایی خصوصیاتی که برای دو کلاس دارای رابطهٔ سلسله‌مراتبی تکرار شده‌اند: در این‌گونه موارد کاربر باید با توجه به شرایط، کلاس پدر یا فرزند را از دامنهٔ خصوصیت مورد نظر حذف کند. به‌عنوان مثال، در زمان ادغام، در دامنهٔ خصوصیت داده «ملیّت» کلاس‌های فرد و دانشجو قرار داشتند. در این مورد به‌طور مشخص، ملیّت باید به‌عنوان خصوصیت کلاس فرد باشد و دانشجو نیازی به این خصوصیت ندارد، زیرا از کلاس پدر آن را به ارث می‌برد؛

۳. شناسایی روابطی که برای دو کلاس در یک رابطهٔ سلسله‌مراتبی تکرار شده‌اند: امکان دارد در زمان ادغام هستان‌نگارها یک رابطه بین کلاس پدر و کلاس فرزند به‌طور مشترک ایجاد شود. به‌عنوان مثال، در مورد هستان‌نگار ساخته‌شده در این پژوهش، رابطهٔ «دریافت‌کردن» بین کلاس فرد و مدرک تحصیلی و همچنین کلاس دانشجو و مدرک تحصیلی به‌وجود آمده بود که به‌دلیل سلسله‌مراتبی بودن کلاس‌های فرد و دانشجو، این رابطه برای کلاس دانشجو حذف شد؛

۴. بررسی پیشنهاد روابط سلسله‌مراتبی: بر اساس مشابهت کلاس‌ها از نظر خصوصیات داده و روابط آن‌ها با سایر کلاس‌ها می‌توان کلاس‌های پدر یا کلاس‌های بالاتر را در سلسله‌مراتب تعریف کرد. به‌عنوان مثال، با اجرای این گام کلاس‌های طرح پژوهشی، گزارش فنی، جزوهٔ درسی، مقاله و کتاب به‌عنوان یک دستهٔ مشابه شناسایی شدند که بر همین اساس، یک کلاس با عنوان سند و با خصوصیات و روابط مشترک این کلاس‌ها ایجاد، و کلاس‌های ذکرشده به‌عنوان فرزندان این کلاس تعریف شدند.

برای ارزیابی هستان‌نگارها از رویکردهای مختلفی استفاده می‌شود که می‌توان آن‌ها

را در دسته‌های ارزیابی بر اساس استاندارد طلایی[1]، بر اساس کاربرد[2]، بر مبنای کاربر[3] و بر اساس استخراج داده[4] قرار داد (Ouyang et al. 2011).

در رویکرد استاندارد طلایی، هستان‌نگار تولیدشده با یک هستان‌نگار مرجع مقایسه می‌شود. از آنجا که چنین استانداردی موجود نبود، ابتدا ساختار اطلاعاتی مورد نیاز فعلی وزارت عتف بررسی شد. در حال حاضر، وزارت عتف داده‌های مورد نیاز خود را از طریق سه سامانهٔ نرم‌افزاری شامل سامانهٔ آموزش عالی[5]، سامانهٔ هوشمند مؤسسات آموزش عالی[6] و سامانهٔ ملی اطلاعات تحقیقاتی[7] تهیه می‌کند. بر اساس قالب فرم‌های دریافت داده در سامانه‌های فوق، یک هستان‌نگار محک[8] ایجاد شد. این هستان‌نگار که ۵۵ کلاس و ۱۵۵ خصوصیت/ رابطه دارد، نشان‌دهندهٔ حداقل مدل داده مورد نیاز است و مقایسهٔ آن با سایر هستان‌نگارها جامعیت هستان‌نگار مقایسه‌شده را نشان می‌دهد. نتایج این مقایسه (جدول ۴) جامعیت بیشتر هستان‌نگار مرجع نسبت به سایر هستان‌نگارها را تأیید می‌کند.

هستان‌نگار محصول این پژوهش در زمان نگارش این مقاله به دفتر فناوری اطلاعات وزارت عتف تحویل شده است تا در طراحی انبارهٔ داده و همچنین، سامانه‌های اطلاعاتی مورد نیاز آن وزارت مورد استفاده قرار گیرد.

جدول ٤. مقایسهٔ هستان‌نگارهای موجود و هستان‌نگار مرجع تولیدی با هستان‌نگار محک

هستان‌نگار	درصد پوشش کلاس‌های محک	درصد پوشش خصوصیات/ روابط محک
هستان‌نگار مرجع	۹۲,۷۳	۹۲,۲۶
وی‌آی‌وی‌اُ	۶۷,۲۷	۲۸,۳۹
سریف	۶۳,۶۴	۲۴,۲۵
اس‌دبلیوآرسی	۳۲,۷۳	۱۰,۹۷
هستان‌نگار تورنتو	۲۱,۸۲	۱۲,۲۶
هستان نگار دانشگاه	۲۱,۸۲	۷,۱
ال‌یوام‌بی	۲۱,۸۲	۳,۸۷
هرو	۲۰	۱۹,۳۵

1. gold standard
2. application based
3. user driven
4. data driven
5. Higher Education System (hes.um.ac.ir)
6. https://portal.irphe.ac.ir/
7. www.semat.ir
8. benchmark

۵. نتیجه‌گیری

یکی از موارد قابل توجه در روش‌های ساخت هستان‌نگار توجه به عناصر هستان‌نگاری موجود است. برای ساخت هستان‌نگارهای بومی باید به موضوع برچسب‌گذاری فارسی عناصر هستان‌نگارهای موجود و چگونگی ادغام آن‌ها و استفاده از دانش پیش‌زمینه (سایر منابع دانشی) توجه شود. با بررسی پیشینهٔ نظری و پژوهشی موضوع، مشخص شد که روش‌های موجود در زمینهٔ ادغام هستان‌نگارها دارای نواقصی همچون عدم به‌روزآوری و خاص‌منظوره بودن، عدم پشتیبانی از زبان فارسی، و دانش پیش‌زمینه بودند. به همین دلیل، پژوهش جاری با هدف ایجاد و آزمون روشی نظام‌مند با جزئیات کافی و پشتیبانی از زبان فارسی برای استفادهٔ مجدد از هستان‌نگارهای موجود در ساخت یک هستان‌نگار بومی انجام شد. برای رسیدن به این هدف، از روش علم طراحی به‌عنوان یک روش کیفی که در مطالعات بین رشته‌ای و به‌خصوص برای تولید راه‌حل‌های عملی کاربرد دارد، استفاده شد. در ادامه، نتایج این پژوهش از نظر رسیدن به هدف اصلی و ویژگی‌های آن بررسی شده است:

۱. داشتن چارچوبی برای فعالیت‌های مورد نیاز تجمیع هستان‌نگارها به قصد استفادهٔ مجدد: بر اساس مطالعهٔ روش‌های پیشین، یک چارچوب اولیهٔ پیشنهادی انتخاب، و در عمل با انجام چرخهٔ روش‌شناسی علم طراحی تغییرات لازم در آن اعمال شد تا در نهایت، به مدل جریان کاری شکل ۲، منجر شد. این چارچوب کلیهٔ فعالیت‌های مورد نیاز را در پنج گام دسته‌بندی کرده و با توجه به استفاده و بهبود آن در عمل، می‌توان ادعا کرد که پوشش کافی را برای فعالیت‌های مورد نیاز فراهم کرده است؛

۲. مشخص‌بودن الگوریتم‌های انجام کار در هر مرحله: روش اجرای هر گام بر اساس مطالعهٔ روش‌های مشابه و تجربهٔ کاری پژوهشگر به‌عنوان پیشنهاد اولیه ارائه و سپس، به‌صورت عملی پیاده‌سازی شد. با اجرای هر الگوریتم میزان کارایی و چالش‌های آن بررسی و در صورت نیاز، تغییراتی در آن اعمال شد تا به کارکرد مطلوب برسد. در شرح این الگوریتم‌ها در بخش یافته‌های این نوشتار به برخی از این چالش‌ها و راه‌حل اعمال‌شده در الگوریتم اشاره شده است. از آنجا که برای پیاده‌سازی یک الگوریتم باید جزئیات آن کاملاً مشخص و فاقد ابهام باشد، می‌توان ادعا کرد که ویژگی بدون ابهام و دارای جزئیات کافی‌بودن، در مورد الگوریتم‌های اجرای هر گام برای روش معرفی‌شده تحقق یافته است.

۳. نقاط ضعف اصلی روش‌های پیشین که در بخش پیشینهٔ نظری این نوشتار شناسایی و معرفی شدند، به‌صورت زیر توسط روش «راما» پوشش داده شده‌اند:

۳-۱. قدیمی‌شدن ابزار و عدم بهبود و توسعهٔ روش‌های ارائه‌شده: ابزار انجام روش «راما» به‌صورت کد منبع باز[1] و با زبان «پی‌اچ‌پی»[2] پیاده‌سازی شده است و به همین دلیل، ارتقای این روش و ابزار آن توسط سایر پژوهشگران به‌سادگی قابل انجام است.

۳-۲. عدم پشتیبانی از زبان فارسی: در «راما» در گام‌های برچسب‌گذاری و نگاشت عناصر از شبکهٔ واژگان فارسی برای مشابهت‌یابی‌های معنایی استفاده شده است.

۳-۳. عدم استفاده از دانش پیش‌زمینه برای بهبود فرایند ساخت: روش «راما» در تعامل با روش «انتیرانداک»، با استفاده از منابع دانشی غیر ساخت‌یافته (مستندات متنی) و مهندسی معکوس پایگاه داده، هستان‌نگار خروجی را غنی‌سازی می‌کند.

از جملهٔ پیشنهادها برای پژوهش‌های آتی در این حوزه می‌توان به ارتقا و بهینه‌سازی الگوریتم ادغام برای در نظر گرفتن قیود هستان‌نگارهای منبع و همچنین، استفاده از هستان‌نگارهای پایه[3] در فرایند تناظربخشی و ادغام اشاره کرد.

فهرست منابع

کمیتهٔ آماربخشی. ۱۳۸۷. *تعاریف و مفاهیم آماری علوم، تحقیقات و فناوری*، تهران: مؤسسه پژوهش و برنامه‌ریزی آموزش عالی.

میلانی‌فرد امید، و محسن کاهانی. ۱۳۹۷. انتیرانداک: سامانهٔ یکپارچهٔ توسعهٔ مشارکتی هستان‌نگار فارسی. *پژوهشنامه پردازش و مدیریت اطلاعات* ۳۳ (۳): ۱۱۲۳-۱۱۵۲.

References

Ba, M., & G. Diallo. 2013. Large-scale biomedical ontology matching with ServOMap. *Innovation and Research in BioMedical Engineering* (*IRBM*) 34 (1): 56-59.

Cheatham, M., & P. Hitzler. 2013. October. String similarity metrics for ontology alignment. In *International Semantic Web Conference*. Berlin, Heidelberg: Springer (pp. 294-309)

Choi, N., I. Y. Song, & H. Han. 2006. A survey on ontology mapping. *ACM Sigmod Record* 35 (3): 34-41.

David, J. 2011. AROMA results for OAEI 2011. In *Proceedings of the 6th International Conference on Ontology Matching-Volume 814*. CEUR-WS. org. (pp. 122-125). Bonn, Germany.

Do, H. H., & E. Rahm. 2002. COMA: a system for flexible combination of schema matching approaches

1. open source 2. PHP 3. foundation ontologies (upper ontologies)

. *In Proceedings of the 28th international conference on Very Large Data Bases*. VLDB Endowment. (pp. 610-621). Hong Kong, SAR, China.

Doan, A., J. Madhavan, P. Domingos, & A. Halevy. 2004. Ontology matching: A machine learning approach. *In Handbook on ontologies*: 385-403. Berlin, Heidelberg: Springer.

Dresch, A., D. P. Lacerda, & J. A. Antunes Jr. 2015. *Design Science Research: A Method for Science and Technology Advancement.* Cham, Switzerland: Springer International Publishing.

Fernández-López, M., & A. Gómez-Pérez. 2002. Overview and analysis of methodologies for building ontologies. *The Knowledge Engineering Review* 17 (02): 129-156.

Gruber, T. R. 1995. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-computer Studies* 43 (5): 907-928.

Guarino, N. 2004. Toward a formal evaluation of ontology quality. *IEEE intelligent Systems* 19 (4): 78-79.

Guarino, N., D. Oberle, & S. Staab. 2009. *What is an Ontology? In Handbook on ontologies*. Berlin Heidelberg: Springer. (pp 1-17).

Huber, J., T. Sztyler, J. Noessner, & C. Meilicke. 2011. CODI: combinatorial optimization for data integration-results for OAEI 2011. In *Proceedings of the 6th International Conference on Ontology Matching-Volume 814*. CEUR-WS. org. (pp. 134-141). Bonn, Germany.

Jain, P., P. Hitzler, A. P. Sheth, K. Verma, & P. Z. Yeh. 2010. Ontology alignment for linked open data. *In International Semantic Web Conference*. Berlin, Heidelberg: Springer. (pp 402-417).

Jiménez-Ruiz, E., B. C. Grau, W. Xia, A. Solimando, X. Chen, V. Cross, & A. Chennai-Thiagarajan. 2014. LogMap family results for OAEI 2014. In *Proceedings of the 9th International Conference on Ontology Matching-Volume 1317*. CEUR-WS. org. (pp 126-134). Aachen, Germany.

Li, J., J. Tang, Y. Li, & Q. Luo. 2008. RiMOM: A dynamic multistrategy ontology alignment framework. *IEEE Transactions on Knowledge and data engineering* 21 (8): 1218-1232.

Madhavan, J., P. A. Bernstein, A. Doan, & A. Halevy. 2005. Corpus-based schema matching. *In Data Engineering, ICDE 2005. Proceedings. 21st International Conference*. IEEE. (pp. 57-68). Tokyo, Japan.

Massmann, S., S. Raunich, D. Aumüller, P. Arnold, & E. Rahm. 2011. Evolution of the COMA match system. *In Proceedings of the 6th International Conference on Ontology Matching-Volume 814*. CEUR-WS. org. (pp. 49-60). Bonn, Germany.

Menéndez-Mora, R. E., & R. Ichise. 2013. Ontology matching by actively propagating user feedbacks through upper ontologies. *Revista Vínculos* 10 (2): 85-92.

Ngo, D., & Z. Bellahsene. 2012. YAM++: A multi-strategy based approach for ontology matching task. In *International Conference on Knowledge Engineering and Knowledge Management*. Berlin, Heidelberg: Springer. (pp 421-425).

Noy, N. F., & D. L. McGuinness. 2001. Ontology development 101: A guide to creating your first ontology. *Stanford Medical Informatics*. Technical Report SMI-2001-0880. Knowledge Systems Laboratory, Stanford University.

Noy, N. F., & M. A. Musen. 2001. August. Anchor-PROMPT: Using non-local context for semantic matching. *In Proceedings of the workshop on ontologies and information sharing at the international joint conference on artificial intelligence (IJCAI)*. (pp 63-70). Seatle, USA.

Otero-Cerdeira, L., F. J. Rodríguez-Martínez, & A. Gómez-Rodríguez. 2015. Ontology matching: A literature review. *Expert Systems with Applications* 42 (2): 949-971.

Ouyang, L., B. Zou, M. Qu, & C. Zhang. 2011. A method of ontology evaluation based on coverage, cohesion and coupling. *In Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on (Vol. 4)*. IEEE. (pp. 2451-2455). Shanghai, China.

Ramar, K., & G. Gurunathan. 2016. Technical review on ontology mapping techniques. *Asian Journal of*

Information Technology 15 (4): 676-688.

Shvaiko, P., & J. Euzenat. 2013. Ontology matching: state of the art and future challenges. *IEEE Transactions on knowledge and data engineering* 25 (1): 158–176.

Suarez-Figueroa, M. C. 2010. NeOn methodology for building ontology networks: Specification, scheduling and reuse . Doctoral dissertation. Madrid: Univ. Politécnica.

Uschold, M., & M. Gruninger. 1996. Ontologies: Principles, methods and applications. *The knowledge engineering review* 11 (02): 93-136.

Vaishnavi, V. K., & W. Kuechler. 2015. *Design science research methods and patterns: innovating information and communication technology.* NW Boca Raton, FL, USA: Crc Press.

W3C. Ontology Dowsing [online]. ESW Wiki. Last modified on 14 January 2010. Available from WWW: http://esw.w3.org/Ontology Dowsing (accessed July 20, 2016)

Zhang, X., Q. Zhong, F. Shi, J. Li, & J. Tang. 2009. Rimom results for oaei 2009. In *Proceedings of the 4th International Conference on Ontology Matching-Volume 551*. CEUR-WS. org. (pp. 208-215). Washington DC, USA.

Zemmouchi-Ghomari, L., & A. R. Ghomari. 2013. Process of building reference ontology for higher education . *In Proceedings of the World Congress on Engineering (Vol. 3)*. (pp 1595-1600). London, UK.

امید میلانی‌فرد

متولد ۱۳۵۶، دارای مدرک دکتری مدیریت فناوری اطلاعات از پژوهشگاه علوم و فناوری اطلاعات ایران است. ایشان هم‌اکنون معاون فنی مرکز فناوری اطلاعات و ارتباطات دانشگاه فردوسی مشهد است. سامانهٔ اطلاعات مدیریت، انبارهٔ داده، مدل‌سازی و مهندسی داده از علایق پژوهشی وی است.

محسن کاهانی

متولد سال ۱۳۴۶، دارای مدرک دکتری مهندسی کامپیوتر از دانشگاه ولونگونگ استرالیاست. ایشان هم‌اکنون استاد تمام گروه مهندسی دانشگاه فردوسی مشهد است.

وب معنایی و فرایندکاوی از جمله علایق پژوهشی وی است.

مهدی علیپور حافظی

متولد سال ۱۳۵۲، دارای مدرک تحصیلی دکتری در رشتهٔ علم اطلاعات و دانش‌شناسی از دانشگاه آزاد اسلامی واحد علوم و تحقیقات است. ایشان هم‌اکنون استادیار دانشگاه علامه طباطبائی است.

حوزه‌های مدیریت منابع اطلاعاتی و کتابخانهٔ دیجیتالی از علائق پژوهشی وی است.