

Ontirandoc: Integrated Collaborative System for Developing Persian Ontology

Omid Milanifard

PhD Candidate in Information Technology Management; Iranian Research Institute for Information Science and Technology (IranDoc);
Corresponding Author milanifard@students.irandoc.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) kahani@um.ac.ir

Received: 11, Dec. 2016 Accepted: 23, Feb. 2017

Abstract: While ontology development is beneficial, it is very costly and time consuming. In order to reduce this cost as well as to increase the accuracy and quality of ontology development, researchers have proposed different methodologies. The goal of these methodologies is to present a systematic manual or semi-automated development of ontologies, while each differs and has its strengths and weaknesses. In this paper, after reviewing current methodologies, we present a new integrated collaborative methodology for ontology development, and compare it with the existing ones. This new system, called Ontirandoc, has been used in two ontology development projects and its accuracy has been evaluated.

Keywords: Ontology, Ontology Development Methodology, Ontology Development System

**Iranian Journal of
Information
Processing and
Management**

**Iranian Research Institute
for Information Science and Technology
(IranDoc)**

ISSN 2251-8223

eISSN 2251-8231

Indexed by SCOPUS, ISC, & LISTA

Vol. 33 | No. 3 | pp. 1111-1140

Spring 2018

<https://doi.org/10.35050/JIPM010.2018.043>



انتیرانداک: سامانه یکپارچه توسعه مشارکتی هستان نگار فارسی

امید میلانی‌فرد

دانشجوی دکتری مدیریت فناوری اطلاعات؛
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛
پدیدآور رابط milanifard@students.irandoc.ir

محسن کاهانی

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



مقاله برای اصلاح به مدت ۲ روز نزد پدیدآوران بوده است.

پذیرش: ۱۳۹۵/۱۱/۲۳

دو یافت: ۱۳۹۵/۰۹/۲۱

فصلنامه | علمی پژوهشی
پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

شاپا (چاپی) ۲۲۰۱-۸۲۲۳

شاپا (الکترونیکی) ۸۲۳۱-۲۲۰۱

نمایه در SCOPUS، و ISC، LISTA

www.irandoc.ac.ir

دوره ۳۳ | شماره ۳ | صص ۱۱۱۱-۱۱۴۰

بهار ۱۳۹۷

<https://doi.org/10.35050/JIPM010.2018.043>

چکیده: ساخت هستان‌نگار با وجود داشتن مزایای بسیار، مستلزم صرف زمان و هزینه زیادی است. محققان در راستای کاهش این هزینه‌ها و همچنین افزایش دقت و کیفیت در فرایند ساخت هستان‌نگار، روش‌های گوناگونی معرفی کرده‌اند. هدف اصلی این روش‌ها، ارائه یک الگوی نظام‌مند دستی یا نیمه‌خودکار برای ساخت هستان‌نگار است. روش‌های موجود با وجود داشتن هدف مشترک، تفاوت‌های زیادی دارند و در هر یک از آن‌ها نقاط قوت و ضعفی وجود دارد. در این مقاله، پس از مروری بر هستان‌نگار و روش‌های ساخت آن، یک سامانه یکپارچه توسعه مشارکتی هستان‌نگار معرفی و ابعاد مختلف آن با سایر روش‌ها مقایسه خواهد شد. از این سامانه جدید با عنوان «انتیرانداک» در دو پروژه جداگانه توسعه هستان‌نگار استفاده و صحت عملکرد آن آزمون شده است.

کلیدواژه‌ها: هستان‌نگار، روش توسعه هستان‌نگار، سامانه توسعه هستان‌نگار



۱. مقدمه

انسان از دیرباز به دنبال شناخت ماهیت و ساختار جهان پیرامون خود بوده است. واژه «آنتولوژی»^۱ را نخستین بار «ارسطو» برای تعریف علم «بودن چون بودن»^۲ در متافیزیک خود به کار برده است. علم آنتولوژی یا هستان‌شناسی زیرشاخه‌ای از فلسفه به حساب می‌آید. با وجود آن که هستان‌شناسی یک علم کاملاً نظری است، کاربردهایی از آن وارد علم اطلاعات شده است. مقصود از آنتولوژی در علم اطلاعات، یک قلم اطلاعاتی است که دانش یک حوزه خاص را مدل‌سازی می‌کند (Guarino, Oberle and Staab 2009). در این تعریف از آنتولوژی، وجود^۳ چیزها و ارتباطات آن‌ها مورد مطالعه قرار می‌گیرد. اما در اینجا «وجود»، یک معنای عملی (و نه فلسفی) دارد: چیزی «وجود» دارد که بتوان آن را بازنمایی کرد (Gruber 1995). از آنجا که در عمل نمی‌توان یک حوزه را به صورت کامل شناخت و مدل اطلاعاتی تنها می‌تواند نگاهی از برخی مفاهیم مهم یک حوزه و روابط بین آن‌ها را بازنمایی کند، استفاده از «نگاشتن» به جای «شناختن» مناسب‌تر به نظر می‌رسد. به همین دلیل، در این نوشتار از اصطلاح «هستان‌نگار» به عنوان معادل فارسی «آنتولوژی» در حوزه علم اطلاعات استفاده شده است.

مشهورترین و پُرارجاع‌ترین تعریف از هستان‌نگار را «گروبر» توصیفی صریح از یک مفهوم‌سازی^۴ معرفی می‌کند (Gruber 1993). «برُست» بخش‌های جدیدی به این تعریف اضافه کرد و هستان‌نگار را «توصیفی رسمی از یک مفهوم مشترک» خواند (Brost 1997). اشتراکی بودن مفهوم به این نکته اشاره دارد که در ساخت مفهوم باید اجماع نظر خبرگان حوزه وجود داشته باشد. همچنین، مفهوم باید به صورت رسمی بازنمایی شود؛ یعنی توسط ماشین قابل خواندن باشد. «ستودر، بنجامینز و فنسل» با ادغام دو تعریف فوق، هستان‌نگار را «توصیفی صریح و رسمی از یک مفهوم‌سازی مشترک»^۵ تعریف کردند (Studer, Benjamins and Fensel 1998).

هدف اصلی ساخت هستان‌نگار، نمایش مدلی انتزاعی از واقعیت است که مورد تأیید افراد ذی‌نفع باشد. واقعیت مدل‌سازی شده می‌تواند تنوع زیادی داشته باشد. از حوزه‌های دانشی بشر گرفته تا سیستم‌های نرم‌افزاری و مخازن اطلاعاتی می‌توانند به عنوان دامنه

1. ontology

2. being qua being

3. existence

4. explicit specification of a conceptualization

5. formal, explicit specification of a shared conceptualization

یا به عبارت دیگر، فضای گفتمان یک هستان نگار باشند. این موضوع باعث پدیدار شدن فضایی وسیع برای کاربرد هستان نگارها شده است. از نظر «آچولد و گرونینگر» سه کاربرد اساسی هستان نگار در بهبود ارتباطات^۱، افزایش قابلیت همکاری^۲ و مهندسی سیستم‌هاست (Uschold and Gruninger 1996). کاربرد هستان نگار در حوزه‌های مختلف علم، فناوری و تجارت همچنان در حال توسعه است. به عنوان مثال، می‌توان از هستان نگار برای مدل‌سازی مفهومی داده (Gailly and Poels 2010) و یا در موتورهای جست‌وجو برای غنی‌سازی متن پرس‌وجوهای کاربران (با اضافه کردن مترادف‌ها و واژگان مرتبط) و در نتیجه، بهبود نتایج جست‌وجو استفاده کرد (Paolucci et al. 2002).

ساخت هستان نگار در عمل، کاری پرهزینه و دشوار است. محققان در راستای کاهش هزینه‌های ساخت و همچنین، افزایش دقت و کیفیت محصول نهایی، روش‌های گوناگونی معرفی کرده‌اند. در هر یک از این روش‌ها سعی شده است از دیدگاهی به فرایند ساخت هستان نگار توجه شده و الگویی نظام‌مند به صورت دستی یا نیمه‌خودکار برای ساخت ارائه داده شود. در این روش‌ها با وجود داشتن هدف مشترک، تفاوت‌های زیادی هست و در آن‌ها به برخی جنبه‌های فرایند ساخت توجه کافی نشده است. همچنین، اغلب آن‌ها دچار کمبود جزئیات در توصیف روش‌ها و فعالیت‌های به کاررفته هستند.

در ادامه مقاله، ابتدا مختصری از پیشینه نظری موضوع بیان شده و در آن برخی از پُراستادترین روش‌های ساخت هستان نگار بررسی و در یک چارچوب با یکدیگر مقایسه شده‌اند. در بخش سوم مقاله، «انتیرانداک» به عنوان یک سامانه یکپارچه ساخت هستان نگار معرفی و در خصوص معماری و پیمانه‌های نرم‌افزاری^۳ آن بحث شده است. در بخش چهارم، پیاده‌سازی و ارزیابی عملی این روش مورد بررسی قرار گرفته و در نهایت، بخش پنجم و انتهای مقاله نتایج حاصل از به کارگیری «انتیرانداک» را بیان می‌کند.

پیشینه نظری

برای ساخت هستان نگارها روش‌های متعددی پیشنهاد شده است. از نظر درجه خودکار بودن می‌توان روش‌های ساخت را به دو دسته غیرخودکار (دستی)^۴ و (نیمه) خودکار^۵ تقسیم کرد. در روش‌های دستی، دانش مفهومی توسط افراد کدگذاری می‌شود

1. communication

2. inter-operability

3. software modules

4. manual

5. (semi) automatic

و در رویکرد نیمه خودکار سعی می‌شود با به کارگیری روش‌ها و الگوریتم‌های مختلف، مداخله انسان در مراحل مختلف تا حد امکان کمتر شده و در نتیجه، سرعت و هزینه ساخت کاهش یابد.

در این بخش ابتدا پُر استنادترین روش‌های ساخت هستان‌نگار به صورت مختصر توضیح داده می‌شود و سپس، مقایسه‌ای بین آن‌ها انجام می‌گیرد.

مروری بر روش‌شناسی‌های موجود برای توسعه هستان‌نگار

سایک: در اواسط دهه ۸۰ میلادی «شرکت MMC»^۲ شروع به ساخت «سایک» به عنوان یک پایگاه عظیم دانش عمومی بشر کرد. برای توسعه آن فرایند زیر دنبال شده است (Lenat et al. 1990):

◇ کدگذاری دستی مقالات و قطعات دانش: این مرحله به صورت دستی انجام شد. چون سامانه‌های موجود پردازش زبان طبیعی، دانش عمومی کافی برای انجام این کار را نداشتند.

◇ کدگذاری دانش با کمک ابزار و با استفاده از دانشی که در مرحله قبل در پایگاه «سایک» قرار گرفته بود.

◇ کدگذاری دانش بر اساس ابزار و با استفاده از دانش موجود در پایگاه «سایک»: در این مرحله، بیشتر کار بر عهده ابزار گذاشته شده است و کاربران تنها منابع را به «سایک» پیشنهاد می‌دهند.

تاکنون از این روش تنها برای توسعه پایگاه دانش «سایک» استفاده شده است.

انترپرایز آنتولوژی: این روش را «آچولد و کینگ» در سال ۱۹۹۵ ارائه دادند و در سال بعد بهینه‌سازی شد (Uschold and Gruninger 1996). بر اساس این روش برای ساخت هستان‌نگار باید فرایندهای زیر انجام شود:

۱. تعیین هدف: شامل تعریف هدف و قلمرو هستان‌نگار بوده و منظور از آن این است که چرا باید هستان‌نگار ساخته شود و کاربردهای مورد انتظار آن و واژگان مرتبط با آن در حوزه مورد نظر چیست.

۲. ساخت هستان نگار: این فرایند شامل سه مرحله دریافت^۱ هستان نگار (شناسایی مفاهیم کلیدی و ارتباط آن‌ها)، کدگذاری مفاهیم و روابط و یکپارچه سازی (استفاده مجدد از هستان نگارهای موجود) است.
۳. ارزیابی: بررسی میزان انطباق هستان نگار ساخته شده با مشخصات هدف گذاری شده آن و قابلیت پاسخگویی هستان نگار به سؤالات شایستگی در این مرحله انجام می شود.
۴. مستندسازی: برای انجام این فرایند راهنماهایی متناسب با هدف و انواع مختلف هستان نگار ارائه شده است.

تاو^۲: این روش که بر اساس تجارب «گرونینگر و فاکس»^۳ در اجرای پروژه «تاو» در سال ۱۹۹۵ پیشنهاد شده است، بیشتر روی فرموله کردن و ارزیابی هستان نگار تمرکز دارد. مراحل پیشنهادی آن عبارت‌اند از:

۱. مشخص کردن مسئله و محدوده هستان نگار: برای این کار مجموعه سناریوهایی که انگیزه ساخت هستان نگار بوده و کاربردهای آن در آینده را مشخص می کنند، نوشته شده و مجموعه راه حل های ممکن برای هر سناریو مشخص می شود. این پاسخ ها ایده اولیه معانی غیررسمی اشیاء و روابط آنهاست؛
۲. ایجاد سؤالات شایستگی غیررسمی بر اساس مرحله اول: سؤالات شایستگی که به زبان طبیعی نوشته می شوند، مشخص می کند که هستان نگار باید به چه سؤالاتی در خصوص سناریوهای نگارش شده مرحله قبل پاسخ دهد. در واقع، نقش این سؤالات، مشخص کردن نیازمندی های هستان نگار است؛
۳. مشخص کردن مجموعه اصطلاحات هستان نگار در زبان منطق مرتبه اول؛
۴. رسمی سازی سؤالات شایستگی (با منطق مرتبه اول)؛
۵. تعریف قواعد و تعاریف برای اصطلاحات (با منطق مرتبه اول)؛
۶. آزمون راه حل ها از نظر کامل بودن در پاسخگویی به سؤالات شایستگی.

مت آنتولوژی: این روش که در سال ۱۹۹۷ توسط «فرناندز، گومز-پرز و جوریستو» پیشنهاد شده، ریشه در فعالیت های اصلی چرخه توسعه نرم افزار دارد و از این نظر بالغ ترین روش

1. capture

2. TOVE (Toronto Virtual Enterprise)

3. Gruning and Fox

4. METHONTOLOGY

موجود است (Fernández-López And Gómez-Pérez & Juristo 2002). این روش شامل فرایند توسعه هستان‌نگار، یک چرخه عمر مبتنی بر الگوهای تکاملی و تکنیک‌هایی برای انجام هر یک از فعالیت‌های مدیریتی، توسعه و پشتیبانی است. ابزارهای مختلفی چون ODE (Blázquez et al. 1998)، WebODE (Corcho et al. 2005)، «پروتزه»^۱ (Noy, Ferguson and Musen 2000) و «آن‌تو ادیت»^۲ (Sure et al. 2002) از این روش پشتیبانی می‌کنند.

فرایند توسعه هستان‌نگار به فعالیت‌هایی اشاره دارد که باید برای ساخت هستان‌نگار انجام شود. توصیه شده است که برای جلوگیری از هرج‌ومرج در ساخت هستان‌نگار سه دسته از فعالیت‌ها انجام گیرد:

- ◇ فعالیت‌های مدیریتی: این فعالیت‌ها شامل برنامه‌ریزی^۳، کنترل، و تضمین کیفیت است. در فعالیت برنامه‌ریزی، کارهایی که باید انجام شود شناسایی شده، سپس، ترتیب انجام و زمان و منابع مورد نیاز برای هر کدام تعیین می‌شود. فعالیت کنترل، اجرای صحیح برنامه‌ریزی را تضمین می‌کند و فعالیت تضمین کیفیت نیز جهت کسب اطمینان از رضایت‌بخش بودن محصول نهایی انجام می‌شود.
- ◇ فعالیت‌های توسعه: این فعالیت‌ها به سه دسته قبل از توسعه، توسعه و پس از توسعه تقسیم می‌شوند. در فعالیت‌های پیش از توسعه، مطالعه محیطی برای شناخت بستر کاربردی هستان‌نگار صورت می‌گیرد. دسته فعالیت‌های زمان توسعه، فعالیت تعریف مشخصات، دلیل ساخت، کاربردها و کاربران هستان‌نگار را مشخص می‌کند. فعالیت مفهوم‌سازی، دانش حوزه را مدل‌سازی می‌کند. فعالیت رسمی‌سازی، مدل مفهوم‌سازی ساخته شده را به مدلی رسمی یا نیمه‌رسمی تبدیل، و در نهایت، فعالیت پیاده‌سازی، مدل را به یک زبان هستان‌نگار پیاده‌سازی می‌کند. مرحله پس از توسعه، شامل فعالیت نگهداری برای به‌روز آوری و تصحیح‌های مورد نیاز بر روی هستان‌نگار است.
- ◇ فعالیت‌های پشتیبانی: شامل کسب دانش (استخراج دانش از خبرگان حوزه یا مستندات موجود توسط فرایندهای نیمه‌خودکار یادگیری هستان‌نگار)، ارزیابی (قضاوت فنی در مورد هستان‌نگار)، یکپارچه‌سازی (استفاده مجدد از سایر هستان‌نگارهای موجود)، مستندسازی (ثبت دقیق جزئیات انجام هر مرحله و محصول آن) و مدیریت پیکره‌بندی (ثبت و ضبط تمام نسخه‌های مستندات و کد هستان‌نگار) است.

در فرایند توسعه هستان نگار، مجموعه فعالیت‌های مورد نیاز جهت ساخت هستان نگار مشخص می‌شود، ولی در خصوص ترتیب انجام کارها صحبتی نمی‌شود. تعیین ترتیب انجام کارها بر عهده چرخه زندگی هستان نگار است. این روش یک چرخه عمر هستان نگار مبتنی بر الگوسازی تکاملی ارائه می‌دهد. به این معنا که در هر بار الگوسازی، نسخه جدیدی از هستان نگار با حذف، اضافه و تغییر واژگان الگوی قبلی ایجاد می‌شود. ساخت هر الگو با برنامه‌ریزی شروع می‌شود. در برنامه‌ریزی باید مشخص شود چه کارهایی و طبق چه زمان‌بندی باید انجام شوند و چه منابعی برای انجام هر کار باید تخصیص داده شوند. پس از برنامه‌ریزی، فعالیت توصیف مشخصات شروع شده و به‌طور همزمان کارهای مختلف مدیریتی (کنترل و تضمین کیفیت) و پشتیبانی (کسب دانش، یکپارچه‌سازی، مستندسازی و مدیریت پیکره‌بندی) آغاز می‌شوند. با ساخته شدن اولین الگو، مدل مفهومی درون فعالیت مفهوم‌سازی ساخته شده است. پس از آن پیاده‌سازی و رسمیت‌بخشی شروع می‌شود. اگر بعد از انجام این فعالیت‌ها نقصانی پیدا شد، می‌توان به عقب برگشت و اصلاحات لازم را انجام داد.

توسعه هستان نگار ۱۰۱: «نوی و مک گینس» این روش را بر اساس این ایده که هیچ راه یگانه‌ای برای مدل‌سازی کردن یک حوزه وجود ندارد، ارائه دادند (Noy and McGuinness, 2001). انتخاب بهترین الگو از بین گزینه‌های موجود، به کاربرد مورد انتظار بستگی دارد. برای توسعه هستان‌نگاری که جهان واقع را به بهترین نحو ممکن منعکس کند، باید هستان نگار را در کاربرد مورد نظر ارزیابی کرد. این ارزیابی‌ها اغلب منجر به اعمال اصلاحات می‌شود. در نتیجه، می‌توان توسعه هستان نگار را فرایندی بازگشت‌شونده دانست. این روش ۷ مرحله را پیشنهاد می‌کند: تعیین حوزه و محدوده، بررسی هستان‌نگارهای موجود جهت استفاده مجدد، شمارش اصطلاحات مهم در هستان نگار (برای این کار تجویز خاصی ندارد)، تعریف کلاس‌ها و سلسله‌مراتب در یک ابزار مدل‌سازی، تعریف خصوصیت‌های هر کلاس، تعریف ویژگی‌ها هر خصوصیت^۲، و ساخت نمونه.

آن-تو-نالج^۳: این روش پیشنهاد می‌کند که برای ساخت هستان نگارها به کاربرد آن‌ها در آینده توجه شود (Sure and Studer 2002)، به همین دلیل هستان‌نگارهای توسعه یافته با این روش وابسته به کاربرد هستند. ویژگی دیگر این روش، تکیه آن بر یادگیری هستان نگار

1. iterative

2. facet

3. slots

4. On-To-Knowledge

برای کاهش زمان و انرژی لازم برای ساخت هستان‌نگار است. فرایندهای پیشنهادی این روش عبارت‌اند از:

۱. مطالعه امکان‌سنجی: لازم است قبل از شروع و به‌عنوان اساس مرحله آغازین^۱ انجام گیرد.
۲. مرحله آغازین: نتیجه این فرایند سند شرح نیازهاست که حوزه، هدف، راهنمای طراحی (به‌عنوان مثال، قواعد نامگذاری)، منابع دانشی موجود (کتاب، مجله، مصاحبه و مانند آن)، کاربران بالقوه و موارد کاربرد هستان‌نگار را مشخص می‌کند. از سؤالات شایستگی می‌توان در این سند استفاده کرد. این سند به سازندگان هستان‌نگار کمک می‌کند در مورد شامل شدن یا نشدن مفاهیم در هستان‌نگار و سلسله‌مراتب مفاهیم تصمیم‌گیری کنند.
۳. پالایش: در این فرایند، پیش‌نویس هستان‌نگار تولیدشده در مرحله قبل در تعامل با خبرگان حوزه کامل می‌شود. پس از تکمیل، هستان‌نگار باید با یک زبان هستان‌نگار پیاده‌سازی شود. ویرایشگر^۲ «آن‌تو ادیت»^۳ برای این کار پیشنهاد شده است.
۴. ارزیابی: در این فرایند با بررسی نیازمندی‌ها و سؤالات شایستگی مشخص می‌شود آیا هستان‌نگار می‌تواند نیازمندی‌های توصیف‌شده را برآورده سازد و به سؤالات شایستگی پاسخ گوید یا خیر. همچنین، هستان‌نگار در محیط کاربرد آن آزمون می‌شود. فرایند پالایش و ارزیابی به‌صورت یک چرخه تا زمان رسیدن به حد مطلوب اجرا خواهند شد.
۵. نگهداری: باید مشخص شود چه کسی مسئول به‌روزآوری و نگهداری است و چگونه این کار انجام می‌شود.

دیلیگنت^۴: «پینتو، تمپیچ و استاب» این روش را با رویکردی کاملاً متفاوت از سایر روش‌های توسعه هستان‌نگار ارائه دادند (Pinto, Tempich and Staab 2009). قابلیت همکاری بین تمام ذی‌نفعان، نقشی اساسی در این روش دارد و ایده اصلی، شروع از یک هستان‌نگار کوچک و قابل استفاده به اشتراک گذاشته شده است. ذی‌نفعان بر اساس اهداف خود، نسخه‌ای

1. kickoff

2. Requirement Specification

3. OntoEdit

4. DILIGENT (Distributed Loosely-controlled evolvInG Engineering of nTologies)

را اقتباس کرده و به صورت محلی استفاده می‌کنند. یک شورای کنترلی از ذی‌نفعان تمام هستان‌نگارهای محلی را بررسی و نسخه جدیدی از هستان‌نگار اشتراکی توسعه خواهد داد. این روش ۵ گام دارد: ۱. ساخت یک هستان‌نگار ساده از ابتدا؛ ۲. تطابق محلی؛ ۳. هستان‌نگار توسط هر یک از ذی‌نفعان به صورت جداگانه استفاده و اختصاصی‌سازی می‌شود؛ ۴. تحلیل: شورای ذی‌نفعان بر اساس درخواست‌های رسیده و مشاهده تغییرات محلی تصمیم می‌گیرد چه تغییراتی بیشترین مزایا را برای کاربران دارد و باید در هستان‌نگار اشتراکی اضافه شود؛ ۵. بازبینی: شورا نسخه بازبینی‌شده از هستان‌نگار را می‌سازد؛ ۶. به‌روزرسانی محلی: کاربران نسخه جدید را در سیستم خود قرار می‌هند. بعد از این مرحله به مرحله ۲ برگشت می‌شود.

اپان: این روش را «نیکلا، میسیکف و نویگلی» با اقتباس از فرایند توسعه نرم‌افزار ارائه داده است (De Nicola, Missikoff and Navigli 2009). فرایند توسعه در یک چرخه جریان کاری^۳ (نیازمندی، تحلیل، طراحی، پیاده‌سازی و آزمون) سازماندهی می‌شود. هر یک از مراحل این چرخه دارای تعدادی گام و دور تکرار است (آغاز، بسط جزئیات، ساخت و گذار) و از مشارکت خبرگان، به‌ویژه در مراحل اولیه توسعه استفاده می‌شود.

نئون: این روش را «سوآرز-فیگورا» در سال ۲۰۱۰ بر پایه روش‌های «دیلیگنت»، «مت‌آنتولوژی» و «آن-تو-نالچ» ارائه داده است (Suarez-Figueroa 2010). این روش چهار پایه اصلی دارد:

۱. واژه‌نامه فرایندها و فعالیت‌ها: به دلیل ماهیت روبه‌تکامل مهندسی هستان‌نگار برخی اصطلاحات در این حوزه ممکن است دارای ابهام باشند. به همین منظور، فهرست واژگانی شامل ۵۹ فعالیت و فرایند به زبان انگلیسی ارائه شده که افراد درگیر پروژه در ارتباط با هم دچار ابهام نشوند.
۲. سناریوهای ساخت هستان‌نگار: بر اساس تجربیات حاصل از انجام چندین پروژه، سه وضعیت متفاوت ساخت هستان‌نگار (هستان‌نگار منفرد، مجموعه هستان‌نگارهای منفرد در تعامل، شبکه‌های هستان‌نگار) تعریف می‌شود. بر همین اساس، ۹ سناریوی مختلف برای ساخت هستان‌نگار معرفی شده که تأکید

1. local adaptation

2. UPON (Unified Process for ONtology)

3. work flow

4. Neon

ویژه آن بر روی استفاده مجدد از منابع دانش موجود و بازمهندسی است.

۳. دو مدل چرخه زندگی هستان‌نگار

۴. مجموعه راهنماهایی برای انجام فرایندها و فعالیت‌ها

مقایسه روش‌ها

چارچوب مقایسه مورد استفاده در این پژوهش شامل میزان پوشش سه دسته اصلی فعالیت مورد نیاز برای ساخت هستان‌نگار (چارچوب مقایسه پیشنهادی، Gómez-Pérez, Fernández-López and Corcho 2004)، پشتیبانی از زبان غیرانگلیسی، ابزار پشتیبانی فنی و میزان بیان جزئیات انجام کار است. جدول شماره ۱، نتیجه این مقایسه را نشان می‌دهد. همان‌گونه که در این جدول مشاهده می‌شود، تمام روش‌های مورد بررسی دارای خلأهایی در پشتیبانی از فرایندهای اصلی توسعه هستان‌نگار هستند و تنها روش ترکیبی «نئون» این پشتیبانی را بر اساس سناریوهای نه‌گانه خود انجام می‌دهد. با وجود این که در بسیاری از موارد، منابع دانشی قابل استفاده در فعالیت اکتساب دانش را متون غیرانگلیسی تشکیل می‌دهند، در این روش‌ها به چگونگی و روند به‌کارگیری و پشتیبانی از چنین منابعی اشاره نشده است.

ابزارهای پشتیبانی فنی از این روش‌ها بیشتر روی فرایند توسعه (به‌خصوص بخش پیاده‌سازی) تمرکز کرده‌اند و برای پشتیبانی فنی از سایر فعالیت‌ها توسعه‌دهندگان باید از ابزارهای دیگر در کنار ابزار اصلی استفاده کنند. به‌عنوان نمونه، هیچ‌یک از این نرم‌افزارها بخشی برای مدیریت پروژه (تعریف و مدیریت اعضای تیم، تخصیص و پیگیری وظایف و منابع)، مدیریت جلسات (برنامه‌ریزی و ثبت جلسات کسب دانش از خبرگان)، پرسشنامه الکترونیکی و سامانه مکاتبات (ارسال و دریافت و مدیریت مستندات تبادل بین اعضای تیم توسعه‌دهنده و خبرگان) ندارند.

جدول ۱. چارچوب مقایسه روش های توسعه هستان نگار

روشناسی	فعالیت های مدیریتی	فعالیت های پیش توسعه	فعالیت های توسعه	فعالیت های پس از توسعه پشتیبانی	فعالیت های حمایت از زبان دیگر	ابزار	داشتن جزئیات
انترپرایز آنتولوژی	-	-	توصیف نیازمندی ها، پیاده سازی	-	کسب دانش، ارزیابی، یکپارچه سازی، مستندسازی	-	خیر
مت آنتولوژی برنامه ریزی، کنترل، تضمین کیفی	-	-	توصیف نیازمندی ها، مفهوم سازی، رسمیت بخشی، پیاده سازی	نگهداری	کسب دانش، ارزیابی، یکپارچه سازی، مدیریت پیکره بندی، مستندسازی	-	پروتژه، بلی آن توادیت
سایک	-	-	پیاده سازی	-	کسب دانش، مستندسازی	-	ابزار خاص خیر
تاو	-	-	توصیف نیازمندی ها، مفهوم سازی، رسمیت بخشی، پیاده سازی	-	کسب دانش، ارزیابی، یکپارچه سازی، مستندسازی	-	خیر
توسعه هستان نگار ۱۰۱	-	-	توصیف نیازمندی ها، مفهوم سازی، رسمیت بخشی، پیاده سازی	-	کسب دانش، یکپارچه سازی	-	خیر
دیلیگت	-	-	-	نگهداری	کسب دانش، ارزیابی، مدیریت پیکره بندی	-	خیر
اپان	-	-	توصیف نیازمندی ها، مفهوم سازی، رسمیت بخشی، پیاده سازی	نگهداری	کسب دانش، ارزیابی	-	خیر

روشناسی	فعالیت‌های مدیریتی	فعالیت‌های پیش‌توسعه	فعالیت‌های توسعه	فعالیت‌های پشتیبانی	حمایت از زبان دیگر	ابزار	داشتن جزئیات
آن-تو-نالچ	برنامه‌ریزی، مطالعه، کنترل، تضمین کیفی	مطالعه، محیطی، امکان‌سنجی	توصیف نیازمندی‌ها، مفهوم‌سازی، رسمیت‌بخشی، پیاده‌سازی	نگهداری، استفاده، ارزیابی، یکپارچه‌سازی، پیکره‌بندی، مستندسازی	کسب دانش، -	آن توادیت خیر	
نئون	برنامه‌ریزی، مطالعه، کنترل، تضمین کیفی	مطالعه، محیطی، امکان‌سنجی	توصیف نیازمندی‌ها، مفهوم‌سازی، رسمیت‌بخشی، پیاده‌سازی	نگهداری، استفاده، ارزیابی، مدیریت، پیکره‌بندی، مستندسازی، ادغام	کسب دانش، بومی‌سازی آن توادیت، بلی پروتزه		

با توجه به نقاط ضعف بیان‌شده در استفاده از روش‌های موجود، در ادامه، سامانه یکپارچه توسعه هستان‌نگار (انتیرانداک) به‌عنوان یک روش جدید توسعه هستان‌نگار معرفی و در خصوص آن بحث می‌شود.

معرفی انتیرانداک

سامانه یکپارچه توسعه هستان‌نگار (انتیرانداک) با الهام از روش‌های موجود، به‌خصوص «آن-تو-نالچ»، «مت‌آنتولوژی» و «نئون» و افزودن ویژگی‌های جدید به‌منظور پوشش نقاط ضعف سایر روش‌ها ایجاد شده است. «انتیرانداک» تنها یک ابزار توسعه نیست، بلکه یک روش کامل شامل توصیف فرایند، فعالیت‌ها و روش‌هایی است که به‌منظور توسعه یکپارچه هستان‌نگار استفاده می‌شوند.

معماری انتیرانداک

معماری «انتیرانداک» در شکل شماره ۱، نشان داده شده است.^۱ این معماری با توجه خاص بر پوشش فعالیت‌های مهم در فرایند ساخت هستان‌نگار و همچنین، امکان انجام مشارکتی این فرایند طراحی شده است. فعالیت‌های مورد نیاز جهت انجام فرایند با آنچه که در سه روش «مت‌آنتولوژی،

۱. این شکل با زبان توصیف معماری «آرچیمیت» از زبان‌های توصیف معماری ذکر شده در استاندارد ISO/IEC/IEEE 42010 ترسیم شده است.

ارتباط بین مفاهیم ایجادشده در هستان‌نگار و مستند نیازها در مرحله ارزیابی هستان‌نگار کاملاً واضح خواهد بود و این موضوع کمک زیادی به انجام دقیق و سریع ارزیابی توسط خبرگان خواهد کرد.

۲. شناسایی منابع دانش: فعالیت‌های مطالعه محیطی و امکان‌سنجی به‌عنوان فعالیت‌های پیش از توسعه، در دو روش «آن-تو-نالچ» و «نئون» مورد اشاره قرار گرفته‌اند. در «انترانداک» این فعالیت‌ها در قالب فعالیت کلی شناسایی منابع دانشی مورد نیاز برای توسعه هستان‌نگار مورد توجه هستند، با انجام این فعالیت مشخص می‌شود چه میزان دانش در قالب‌های مختلف برای ساخت هستان‌نگار در دسترس است و اصولاً نیازی به ساخت هستان‌نگار جدید وجود دارد یا خیر. برای انجام این فعالیت، اجرای موازی چند فعالیت پیشنهاد می‌شود:

۱-۲. شناسایی هستان‌نگارهای مرتبط موجود: استفاده مجدد از هستان‌نگارهایی که توسط توسعه‌دهندگان دیگر برای حوزه موضوعی تولید شده‌اند، می‌تواند به میزان قابل توجهی در سرعت ایجاد هستان‌نگار و دقت آن مؤثر باشد. این امکان نیز وجود دارد که هستان‌نگاری یافت شود که تا حد قابل قبولی هدف ایجاد هستان‌نگار را پوشش دهد و در نتیجه، نیاز به ادامه کار وجود نداشته باشد. پس از انجام این فعالیت، انجام فعالیت ثبت و پالایش اطلاعات هستان‌نگارها پیشنهاد شده است که با استفاده از دو پیمانۀ نرم‌افزاری تبدیل کد هستان‌نگار به عناصر هستان‌نگاری و برچسب‌گذاری عناصر هستان‌نگاری انجام می‌شود. پیمانۀ تبدیل کد، فایل هستان‌نگار پیداشده را به زبان ¹OWL دریافت و پس از جداسازی عناصر اصلی هستان‌نگاری (کلاس‌ها و خصوصیات) آن‌ها را در مخزن عناصر هستان‌نگاری ذخیره می‌کند. توسعه‌دهندگان هستان‌نگار می‌توانند با استفاده از پیمانۀ برچسب‌گذاری، به هر یک از عناصر هستان‌نگار یک برچسب معنایی به‌زبان فارسی نسبت دهند. این پیمانۀ که قابلیت شناسایی و پیشنهاد برچسب بر اساس مشابهت ساختاری و معنایی را دارد، معادلی برای فعالیت بومی‌سازی است که تنها در روش «نئون» به آن اشاره شده است.

1. Web Ontology Language (WOL)

برچسب گذاری عناصر به زبان فارسی که لایه معنایی جدیدی برای عناصر هستان نگاری فراهم می کند، یکی از مزایای قابل توجه «انتیرانداک» نسبت به سایر روش ها برای استفاده در ساخت هستان نگارهای بومی است.

۲-۲. شناسایی مستندات غیرساخت یافته مرتبط با حوزه: مستندات غیرساخت یافته (متنی) معمولاً به عنوان منبع اصلی دانش حوزه شناخته می شود. تنوع اسناد قابل استفاده ممکن است بسیار گسترده باشد. در این فعالیت لازم است اسناد از نظر میزان دانش و اعتبار آنها اولویت بندی شوند (به عنوان مثال، اصطلاح نامه های حوزه و همچنین، اسناد بالادستی سازمانی مانند سند برنامه راهبردی و شاخص های ارزیابی حوزه کسب و کار از معتبرترین و مهم ترین اسناد دانشی هستند). برای انجام این فعالیت استفاده از سه پیمانۀ نرم افزاری پیشنهاد می شود. توسعه دهندگان پس از یافتن هر مستند با استفاده از پیمانۀ ثبت مستندات آن را در یک مخزن ذخیره سازی می کنند. این پیمانۀ علاوه بر ثبت فایل، امکان ثبت محتوای هر صفحه را به صورت مجزا فراهم می کند. استخراج محتوای مستند و ذخیره سازی آن در صفحات مجزا امکان اعمال روش های استخراج نیمه خودکار واژگان از متن با روش های آماری و زبان شناختی (در پیمانۀ استخراج نیمه خودکار واژگان) و انجام تحلیل هم رخدادی برای تشخیص روابط معنایی بین واژگان استخراج شده (در پیمانۀ نرم افزاری تشخیص روابط و ارجاعات واژگان) و همچنین، نمایش منبع هر عنصر مدل سازی شده در هستان نگار نهایی را (در پیمانۀ دریافت نظرات خبرگان) برای افزایش دقت و صحت عمل ارزیابی هستان نگار فراهم می آورد.

۳-۲. شناسایی پایگاه داده های عملیاتی حوزه: ساختار و محتوای پایگاه داده های عملیاتی موجود می تواند منبع خوبی برای استخراج مفاهیم و روابط میان آنها در حوزه مورد نظر باشد. در این فعالیت باید تلاش شود سامانه های عملیاتی که حداکثر پوشش خود کارسازی عملیات را در حوزه مورد نظر دارند (به عنوان مثال، سامانه های برنامه ریزی منابع سازمان^۱ در حوزه

1. ERP (Enterprise Resource Planning)

مورد نظر) شناسایی، و دسترس‌پذیری به پایگاه داده‌های آن‌ها برای انجام مهندسی معکوس مشخص شود. طرحواره^۱ پایگاه داده‌های مورد نظر با استفاده از پیمانۀ انتقال طرحواره پایگاه داده به مخزن فراداده به‌عنوان نسخه اولیه از ساختار پایگاه داده ذخیره‌سازی می‌شود. از آنجا که معمولاً طرحواره پایگاه داده شامل فراداده کامل برای مهندسی معکوس نیست، فعالیت بعدی با عنوان تکمیل دستی مستندات بخشی از روابط بین عناصر داده در طرحواره به شکل کلیدهای خارجی و اصلی وجود نداشته و پس از این کار لازم است فعالیت تکمیل فراداده اجرا شود. در این فعالیت اطلاعات تکمیلی در مورد ساختار پایگاه داده توسط پیمانۀ نرم‌افزاری تکمیل دستی مستندات (شرح فیلدها و جداول اطلاعاتی)/شناسایی اتصالات (شناسایی ارتباطات بین جداول که در طرحواره به‌شکل کلید خارجی وجود ندارند) ثبت می‌شود.

۴-۲. شناسایی خبرگان حوزه: خبرگان، یک منبع دانشی ارزشمند برای شناخت حوزه هستند و استفاده از نظرات آن‌ها جهت ارزیابی و اصلاح هستان‌نگار باعث اعتباربخشی به محصول نهایی خواهد شد. در این فعالیت باید فهرستی از خبرگان به همراه اولویت‌بندی آن‌ها از نظر میزان دانش و اعتبار و همچنین، میزان دسترس‌پذیری به آن‌ها مشخص شود. چنانچه این فهرست اعضای کمی داشته باشد، توسعه و ارزیابی هستان‌نگار دشواری بیشتری خواهد داشت و حتی ممکن است مقرون‌به‌صرفه تشخیص داده نشود. اطلاعات خبرگان با کمک پیمانۀ نرم‌افزاری ثبت اطلاعات خبرگان ثبت و ضبط می‌شود. خبرگان ثبت شده در سامانه می‌توانند از طریق پیمانۀ نرم‌افزاری دریافت نظرات خبرگان محصول نهایی را مشاهده و نظرات اصلاحی خود را ثبت کنند.

۳. مهندسی معکوس پایگاه داده: این فعالیت به‌همراه فعالیت‌های استخراج واژگان از متن و ادغام هستان‌نگارهای موجود، سه فعالیت هستند که در راستای استخراج دانش از منابع شناسایی شده در فعالیت قبلی انجام می‌شوند. فعالیت مهندسی

1. schema

معکوس پایگاه داده، زمانی قابل انجام است که پایگاه داده مناسبی در حوزه موضوعی پیدا شده و فراداده آن تکمیل شده باشد. خروجی این فعالیت که به کمک پیمانۀ نرم‌افزاری مهندسی معکوس انجام می‌شود، هستان‌نگاری است که عناصر آن به مخزن عناصر هستان‌نگاری موجود در حوزه موضوعی اضافه می‌شود. از این عناصر در فعالیت مفهوم‌سازی استفاده خواهد شد. روش پیشنهادی برای انجام مهندسی معکوس بر پایه وجود یک فراداده تکمیل شده است که شامل شرح جداول و فیلدها و مشخص کردن ارتباطات ساختاری (کلیدهای خارجی در سطح پایگاه داده) و معنایی (ارتباط جداول در سطح برنامه کاربردی) میان جداول است.

۴. استخراج واژگان: همان‌طور که اشاره شد، منابع غیرساخت‌یافته مهم‌ترین منابع دانشی حوزه هستند. در این فعالیت، واژگان مورد استفاده در مستندات غیرساخت‌یافته، که به صورت صفحه‌های مجزا جداسازی شده‌اند، توسط اعضای تیم توسعه و با کمک پیمانۀ نرم‌افزاری استخراج نیمه‌خودکار واژگان به دست آمده و پس از پالایش و ادغام (با کمک پیمانۀ نرم‌افزاری شناسایی واژگان ساختاری و معنایی) در مخزن واژگان ثبت و ضبط می‌شوند.

۵. تحلیل و انتخاب هستان‌نگارهای موجود برای ادغام: در این فعالیت، هستان‌نگارهای یافت‌شده در مرحله قبل با کمک پیمانۀ نرم‌افزاری تحلیل فراوانی و میزان همپوشانی مفاهیم هستان‌نگارها بررسی شده و بر اساس اطلاعات آماری حاصل و همچنین میزان تناسب مفاهیم آن‌ها با حوزه مورد نظر، در خصوص انتخاب هستان‌نگارهای مناسب برای ادغام، تصمیم‌گیری می‌شود.

۶. ادغام هستان‌نگارها: این فعالیت پس از انجام تحلیل و انتخاب هستان‌نگارهای موجود برای ادغام قابل انجام است. در این فعالیت مجموعه هستان‌نگارهای منتخب با استفاده از پیمانۀ نرم‌افزاری پالایش عناصر هستان‌نگاری با هم ادغام می‌شوند و عناصر هستان‌نگاری حاصل در قالب یک هستان‌نگار جدید به مخزن افزوده می‌شود.

۷. مفهوم‌سازی واژگان: در این فعالیت توسعه‌دهندگان با بررسی فهرست واژگان مستخرج با کمک پیمانۀ نرم‌افزاری استفاده مجدد از عناصر هستان‌نگاری و تشخیص روابط و ارجاعات واژگان در خصوص مدل‌سازی کردن هر واژه به یک

عنصر هستان‌نگار، تصمیم‌گیری کرده و عنصر مدل‌سازی شده را با استفاده از پیمانۀ نرم‌افزاری نگاشت واژگان به عناصر هستان‌نگاری در مخزن ثبت می‌کنند. ۸ ارزیابی خبرگان: پس از انجام فعالیت‌های توسعه، هستان‌نگار هدف شکل گرفته است و باید صحت و دقت آن مورد بررسی قرار گیرد. در این فعالیت، خبرگان با استفاده از پیمانۀ نرم‌افزاری دریافت نظرات خبرگان، در خصوص تمامی اجزای هستان‌نگار تولید شده اظهار نظر می‌کنند. این نظرات در اختیار تیم توسعه قرار می‌گیرد تا طی فعالیت بعدی برای تصحیح و بهبود هستان‌نگار مورد استفاده قرار گیرد.

۹. نگهداری و ارتقاء: پس از ساخت نسخه اولیه هستان‌نگار لازم است بر اساس اطلاعات حاصل از ارزیابی خبرگان، تیم توسعه‌دهنده تغییرات لازم را در هستان‌نگار اعمال کند. این کار ممکن است چندین بار و به صورت متناوب تا رسیدن به نتیجه دلخواه تکرار شود. در نهایت، با توجه به پویایی محیط، هستان‌نگار نیز باید به صورت مداوم تحت بازبینی و ارتقاء باشد تا مدل قابل قبولی از واقعیت باقی بماند.

۱۰. مدیریت پروژه: توسعه یک هستان‌نگار با توجه به میزان زمان و انرژی لازم، اغلب نیازمند مشارکت چندین نفر در قالب یک تیم توسعه‌دهنده است. «انتیرانداک» پیشنهاد می‌کند که فعالیت مدیریت پروژه به‌عنوان یک فعالیت موازی و جانبی در کنار سایر فعالیت‌های ذکر شده انجام شود. پیمانۀ نرم‌افزاری مکاتبات الکترونیکی و مدیریت کارها و مستندات، به تیم مشارکت‌کننده در انجام کار کمک خواهد کرد.

مهم^۱ است. هر یک از اعضای تیم، نام کاربری و کلمه عبوری دریافت خواهند کرد که از طریق آن می‌توانند با ورود به سایت سامانه کارتابل کار خود را مشاهده کرده و شرح دقیق کاری را که باید انجام دهند، مشاهده کنند. همچنین، می‌توانند نتایج کاری خود را تحت عنوان اقدامات به هر کار پیوست کنند. زمان‌های انجام کار به صورت تجمیع شده در گزارشات مدیریتی به مدیر پروژه نمایش داده خواهد شد و با توجه به ثبت قراردادهای کاری افراد در سامانه، امکان ثبت و ضبط امور مالی و پرداختی به اعضای تیم را فراهم می‌آورد. همچنین، کلیه فعالیت‌هایی که افراد در سامانه انجام می‌دهند، در مخزن ممیزی عملیات ذخیره شده و در نتیجه، مستندسازی فعالیت‌ها در سطح ریزترین جزء به صورت خودکار انجام خواهد شد.

۲. مکاتبات الکترونیکی: این پیمانانه امکان تبادل الکترونیکی متون و مستندات را بین اعضای پروژه فراهم می‌آورد. از سوابق پیام‌ها و فایل‌های رد و بدل شده در این پیمانانه می‌توان در مستندسازی کار استفاده کرد.

۳. مدیریت جلسات: این پیمانانه امکان تعریف و درخواست حضور در جلسات، تعیین دستور کار، ثبت و ضبط صورت‌جلسه‌ها و اقدامات و یادداشت‌های مرتبط با هر جلسه را فراهم می‌آورد. جلسات برگزار شده می‌تواند مربوط به برگزاری مصاحبه با خبرگان برای استخراج دانش باشد که در این صورت، از نتایج جلسه (فایل مستندات متنی) به عنوان منبع دانشی غیرساخت یافته استفاده خواهد شد و ممکن است جلسه مربوط به هماهنگی و امور اجرایی تیم توسعه‌دهنده باشد که در این صورت، نتیجه جلسه به مستندسازی پروژه کمک خواهد کرد.

۴. پرسشنامه الکترونیکی: با استفاده از این پیمانانه اعضای تیم می‌توانند پرسشنامه طراحی کرده و سپس، برای جامعه آماری مورد نظر خود حساب‌های کاربری موقت برای پر کردن پرسشنامه ایجاد کنند. پرسش‌شوندگان می‌توانند از طریق وب، اطلاعات پرسشنامه را تکمیل کنند. از این پیمانانه می‌توان برای استخراج دانش از خبرگانی که دسترسی حضوری به آن‌ها مقدور نیست، استفاده کرد. ۵. ثبت اطلاعات خبرگان: این پیمانانه امکان مدیریت فهرست خبرگان در دسترس را

1. Milestone (تاریخ‌هایی که باید خروجی مشخصی از پروژه گرفته شود)

فراهم می‌آورد.

۶. مدیریت هستان نگارها: هستان نگارهایی که در طی فعالیت جست‌وجوی منابع دانشی پیدا شده‌اند و یا توسط فعالیت‌های ادغام هستان نگار، مفهوم‌سازی واژگان و یا مهندسی معکوس ایجاد می‌شوند، توسط این زیرسامانه ثبت و مدیریت می‌شوند. پیمان‌های مختلفی که در این زیرسامانه استفاده می‌شوند، عبارت‌اند از:

◇ تبدیل کد هستان نگار به عناصر هستان نگاری: این پیمان‌ها با دریافت کد هستان نگارهای موجود به زبان OWL آن‌ها را به عناصر هستان نگاری، تجزیه و در مخزنی ثبت می‌کند. مدیریت فراداده مربوط به هستان نگارها (عنوان، مسیر فایل در وب و ...) از دیگر امکانات این پیمان‌ها است.

◇ برچسب‌گذاری عناصر هستان نگاری: این پیمان‌ها امکاناتی فراهم می‌کند که کاربران بتوانند به هر یک از عناصر هستان نگارها، یک برچسب فارسی نسبت دهند. در این پیمان‌ها نیز از الگوریتم‌های مشابهت‌یابی ساختار و معنایی برای پیدا کردن برچسب‌های مشابه و پیشنهاد ادغام آن‌ها به کاربر استفاده می‌شود. همچنین، این پیمان‌ها بر اساس اطلاعاتی که از پیش موجود است، به کاربر برچسب‌هایی برای عناصر هستان نگاری پیشنهاد می‌دهد.

◇ تحلیل فراوانی و میزان پوشش هستان نگارها: این پیمان‌ها با استفاده از لایه معنایی جدید حاصل از برچسب‌گذاری فارسی عناصر هستان نگارها، آن‌ها را از نظر میزان همپوشانی مفاهیم و خصوصیات با یکدیگر مقایسه می‌کند. بر اساس نتایج حاصل از این بررسی، تیم توسعه‌دهنده هستان نگارهایی را برای ادغام، انتخاب خواهد کرد تا از عناصر آن‌ها در هستان نگار نهایی استفاده شود.

◇ پالایش (ادغام و حذف) عناصر هستان نگاری: این پیمان‌ها امکان تعریف پروژۀ ادغام هستان نگارها را فراهم می‌آورد. در یک پروژۀ ادغام، چند هستان نگار منتخب با استفاده از الگوریتم ادغام نیمه‌خودکار «پرامپت»^۱ (Noy and Musen 2003) با هم ادغام می‌شوند. همچنین، این پیمان‌ها امکان ادغام عناصر موجود

در یک هستان‌نگار را هم فراهم می‌سازد. پیشنهادات ادغام ارائه‌شده توسط این پیمان‌ه بر اساس مشابهت‌های ساختاری و معنایی نام اصلی عناصر (به زبان انگلیسی) و برچسب‌های معنایی ثبت‌شده (به زبان فارسی) است. استفاده از دو زبان در فرایند ادغام از ویژگی‌های منحصر به فرد «انترانداک» است.

◇ دریافت نظرات خبرگان (ارزیابی هستان‌نگار): برای ارزیابی، پیمانۀ دریافت نظرات خبرگان امکان منحصر به فردی را جهت تبدیل هستان‌نگار به مجموعه گزاره‌های فارسی جهت پرسش از خبرگان به صورت الکترونیکی فراهم می‌کند. برای انجام این کار باید از قواعد نام‌گذاری مشخصی در زمان برچسب‌گذاری فارسی مفاهیم هستان‌نگار پیروی کرد: کلیۀ مفاهیم با نام مفرد و روابط معنایی بین مفاهیم (خصوصیات شیء) به شکل مصدر فعل برچسب‌گذاری شوند. پس از پایان عملیات مفهوم‌سازی، مدیر پروژه خبرگانی را برای ارزیابی به هستان‌نگار منتسب می‌کند. این خبرگان می‌توانند از طریق آدرسی که در اختیار آن‌ها قرار گرفته با ثبت یک کد رمز وارد سامانه شده و به ترتیب، برای تمام مفاهیم، خصوصیات و روابط مدل‌سازی شده اظهار نظر کنند. اظهار نظر آن‌ها به شکل پذیرش، رد یا نداشتن نظر روی گزاره‌هایی خواهد بود که به زبان فارسی به آن‌ها نشان داده می‌شود. همچنین، خبرگان می‌توانند نظرات خود را در خصوص هر یک از عناصر هستان‌نگاری به صورت متنی در سامانه ثبت کنند. مدیر پروژه با مشاهده تجمیع نظرات، آن‌ها را برای اعمال در هستان‌نگار به افراد ذی‌ربط منتسب خواهد کرد.

7. مدیریت منابع دانشی غیرساخت یافته: این زیرسامانه برای کمک به توسعه‌دهندگان برای ثبت منابع غیرساخت یافته و انجام عملیات مختلف استخراج واژگان و مدل‌سازی آن‌هاست. پیمان‌ه‌های نرم‌افزاری این زیرسامانه عبارت‌اند از:

◇ ثبت مستندات و تجزیۀ محتوا به صفحات: با استفاده از این پیمان‌ه می‌توان فایل و اطلاعات تکمیلی منابع غیرساخت یافته و محتوای آن‌ها را به صورت صفحه‌بندی شده ذخیره‌سازی و پایش کرد. همچنین، این پیمان‌ه با فراهم آوردن گزارش آماری از میزان واژگان استخراج شده از هر صفحه، امکان نظارت و کنترل بر فرایند استخراج واژگان را فراهم می‌آورد.

- ◇ استخراج نیمه خودکار واژگان از متن: این پیمانانه امکان شناسایی واژگان مهم هر صفحه را با استفاده از روش‌های زبان‌شناختی فراهم می‌آورد. فهرست واژگان پیشنهادی به کاربر نمایش داده می‌شود تا از بین آن‌ها موارد صحیح و مرتبط را انتخاب و به مخزن اضافه کند. در زمان استخراج واژگان، توسعه‌دهندگان باید شماره صفحه و پاراگرافی را که واژه در آن مورد ارجاع قرار گرفته است، ثبت کنند. با این کار در مرحله مفهوم‌سازی، امکان بررسی واژگان هم‌رخداد با تحلیل هم‌رخدادی امکان‌پذیر خواهد شد.
- ◇ شناسایی واژگان مشابه ساختاری و معنایی و ادغام آن‌ها: این پیمانانه با کمک روش مشابهت‌یابی ساختاری (فاصله لونشتین)^۱ و معنایی (استفاده از شبکه واژگان فارسی)، به کاربران کمک می‌کند که واژگان مستخرج از متون مشابه را با هم ادغام کنند. هرگونه ادغام یا حذف واژگان در سوابق ثبت می‌شود.
- ◇ مفهوم‌سازی واژگان: با استفاده از این پیمانانه، توسعه‌دهندگان واژگانی را که از متون استخراج شده، به صورت یک فهرست مرتبط بر اساس میزان اهمیت (با استفاده از روش وزن‌دهی فراوانی واژه و وزن‌دهی فراوانی واژه-معکوس فراوانی سند^۲) مشاهده کرده و با انتخاب هر واژه می‌توانند آن را به یکی از عناصر هستان‌نگاری مدل‌سازی کنند. این پیمانانه برای مفهوم‌سازی، امکانات گوناگونی را فراهم می‌آورد. از جمله آن‌ها می‌توان به موارد زیر اشاره کرد:
 - تحلیل معنایی واژه با استفاده از شبکه واژگان فارسی «فردوس‌نت» و نمایش واژگان هم‌رده، عام‌تر و خاص‌تر (برای تشخیص و ساخت سلسله‌مراتب مفاهیم)
 - تحلیل ساختار نحوی (نمایش واژگان مشابه از نظر فاصله «لونشتین»، واژگان مشابه با پسوند یا پیشوند و همچنین، واژگان حاوی این واژه) برای تشخیص ارتباطات معنایی (خصوصیات شیء در هستان‌نگار)
 - نمایش عناصر مشابه موجود در مخزن هستان‌نگار (حاصل از انجام پروژه‌های ادغام) بر اساس برچسب‌گذاری فارسی انجام‌شده (توسعه‌دهنده

1. Levenshtein distance

2. TF/IDF (Term Frequency/ Inverse Document Frequency)

از این طریق می‌تواند با دیدگاه سایر مدل‌سازان نسبت به آن واژه در هستان‌نگارهای موجود آشنا شده و انتخاب بهتری برای مفهوم‌سازی انجام دهد)

□ مشاهده‌ی واژگان هم‌رخداد و فراوانی هم‌رخدادی (بر این اساس می‌توان روابط معنایی بین مفاهیم را مدل‌سازی کرد).

۸ مدیریت فراداده پایگاه داده و مهندسی معکوس: این زیرسامانه برای استخراج عناصر هستان‌نگاری از پایگاه داده به تیم توسعه کمک می‌کند و شامل پیمانانه‌های نرم‌افزاری زیر است:

◇ انتقال طر حواره پایگاه داده به مخزن فراداده: این پیمانانه با اتصال به موتور پایگاه داده، طر حواره آن را دریافت و به مخزن فراداده «انتیرانداک» اضافه می‌کند. چنانچه عناصر طر حواره (جداول و فیلدها) دارای شرح و توضیح باشند، به صورت پیش فرض به عنوان برچسب آن عنصر در فراداده ثبت می‌شوند.

◇ پیمانانه تکمیل دستی مستندات / شناسایی اتصالات: روش پیشنهادی «انتیرانداک» برای مهندسی معکوس پایگاه داده بر پایه ایجاد یک فراداده مناسب پایگاه داده عمل می‌کند. فراداده‌ای مناسب است که اجزای پایگاه داده (شامل جداول، فیلدها، شرح مقادیر برای فیلدهای شمارشی و اتصالات جداول خارج از طر حواره پایگاه داده) به صورت دستی یا نیمه خودکار مستند شده باشند.

◇ مهندسی معکوس: این پیمانانه بر اساس قوانین نگاشت عناصر پایگاه داده رابطه‌ای به عناصر هستان‌نگاری، یک هستان‌نگار از روی فراداده ایجاد کرده و عناصر آن را به مخزن عناصر هستان‌نگاری اضافه می‌کند. الگوریتم مورد استفاده اقتباسی از الگوریتم (Zaremba 2015) است.

پیااده‌سازی و ارزیابی عملی:

از «انتیرانداک» برای اجرای دو پروژه ساخت هستان‌نگار بومی (هستان‌نگار مرجع مؤسسات آموزشی و پژوهشی زیرمجموعه «وزارت علوم، تحقیقات و فناوری»، به سفارش

«دفتر فناوری اطلاعات وزارت عتف»^۱، و هستان نگار پژوهش و فناوری استان «خراسان رضوی» به سفارش «منطقه ویژه علم و فناوری رضوی»، استفاده شده و بر اساس تجربیات حاصل بهبود یافته است. در ادامه، شرح مختصری از ایجاد هستان نگار پژوهش و فناوری استان «خراسان رضوی» به عنوان کاربرد عملی این روش آمده است. در شروع کار، تیم توسعه توسط مدیریت پروژه انتخاب گردید و ثبت و دسترسی های آن‌ها در پیمانۀ مدیریت پروژه تنظیم شد.

مدیر پروژه در طول فرایند توسعه، با استفاده از پیمانۀ مدیریت پروژه، عملیات مختلف مدیریتی از جمله تخصیص کار به اعضا، مدیریت مستندات، نظارت بر انجام کارها و اقدامات و اخذ گزارشات مختلف زمانی و مالی را انجام می‌داد. بخش مکاتبات الکترونیکی و مدیریت جلسات نیز به عنوان ابزار حمایتی از روند کلی انجام کار توسط اعضا مورد استفاده قرار می‌گرفتند.

پس از انجام فعالیت‌های شناسایی منابع دانش، هستان نگارهای یافت‌شده (۱۰ مورد) در بخش مدیریت هستان نگارها (شکل ۳) و مستندات مرتبط (۲۸ مورد) نیز در بخش مدیریت منابع استخراج واژگان وارد شدند.

ردیف	عنوان	فهرست‌دهی	فایل	شرح	تاریخ	تعداد بازدید	تعداد کامنت	تعداد اشتراک
1	FOAF		http://xmlns.com/foaf/spec/index.rdf	هستان نگار انحصاری و شبکه اجتماعی آنها	1395/09/15	15	66	3
2	LUMB		http://swat.cse.lehigh.edu/onto/univ-bench.owl	Lehigh University Benchmark	1395/09/15	43	31	3
3	SWRC		http://swrc.ontoware.org/ontology	وب‌های برای انحصاری پژوهشی	1395/09/15	55	70	3
4	DBLP		http://swat.cse.lehigh.edu/resources/onto/db.owl	هستان نگار پژوهشی	1395/09/15	22	23	4
5	ro		http://w4ever.github.com/ro.owl	هستان نگار پژوهشی	1395/09/15	15	12	5
6	FRAPo		http://eest.cs.umbo.it/apps/LODE/source?url=http://purl.org/cerif/frapo	هستان نگار سرچشمه گزارش، مدیریت پژوهشی و پروژه‌ها	1395/09/15	94	138	6
7	scoro		http://eest.cs.umbo.it/apps/LODE/source?url=http://purl.org/lsqr/scoro	هستان نگار نقشه و مشارکت‌های علمی	1395/09/15	35	78	7
8	VIVO		http://vivoweb.org/files/voaf-public-1.6.owl	هستان نگار حوزه ی آموزشی و پژوهشی	1395/09/15	35	78	8

شکل ۳. مدیریت هستان نگارهای موجود

عناصر هستان‌نگارهای یافت‌شده (کلاس‌ها و خصوصیات)، با استفاده از امکانات شرح داده‌شده در ۷-۲ برچسب‌گذاری فارسی و ادغام شدند (شکل ۴). بر اساس عناوین عناصر و برچسب‌های آن‌ها از نظر میزان همپوشانی مفاهیم و روابط تحلیل آماری انجام شده و در نهایت، ۵ هستان‌نگار برای ادغام، انتخاب و با کمک ابزار ادغام شدند که حاصل آن هستان‌نگاری با ۱۰۲ کلاس و ۲۷۹ خصوصیت داده و شیء بود.

عنوان کلاس		برچسب		کلاس بالاتر	
1	College	آموزشگاه	[ورایش]	Organization	[ورایش]
2	Lecturer	آموزشگر	[ورایش]	Faculty	[ورایش]
3	Professor	استاد	[ورایش]	Faculty	OBJECT: advisor (راهنما / مشاور) OBJECT: tenured (استفاد داشتن)
4	FuProfessor	استاد تمام	[ورایش]	Professor	
5	VisitingProfessor	استاد مدعو	[ورایش]	Professor	
6	AssatantProfessor	استادیار	[ورایش]	Professor	
7	Department	بخش	[ورایش]	Organization	
8	Program	برنامه	[ورایش]	Organization	
9	Research	پژوهش	[ورایش]	Work	OBJECT: publicationResearch (نگارش پژوهشی) OBJECT: researchProject (پروژه پژوهشی)
10	PostDoc	پسا دکتری	[ورایش]	Faculty	
11	Student	دانشجو	[ورایش]		UndergraduateStudent (دانشجوی کارشناسی) [ورایش]

شکل ۴. برچسب‌گذاری عناصر هستان‌نگار

به صورت موازی، کار استخراج واژگان از منابع غیرساخت یافته توسط اعضای تیم صورت گرفت و در نتیجه، ۶۶۲ واژه به دست آمد. این ابزار در فرایند استخراج واژگان، حمایت‌هایی چون پیشنهاد واژه‌های مشابه ساختاری و معنایی، امکان ادغام واژگان مشابه در حین کار و نمایش عنصر مدل‌سازی شده در هستان‌نگار برای اعضای تیم فراهم می‌کرد. پس از استخراج واژگان، عملیات نگاشت هر واژه به عنصر هستان‌نگاری معادل، با کمک ابزاری که تحلیل‌های مختلف مشابهت‌یابی ساختاری و معنایی، استفاده از عناصر هستان‌نگاری موجود و همچنین، تحلیل هم‌رخدادی انجام می‌داد، صورت گرفت (شکل ۵).

یادشده در مرحله ارزیابی نهایی است.

نتیجه‌گیری

روش «انتیرانداک»، فرایند توسعه ساخت هستان‌نگار را به‌صورت مجموعه‌ای از فعالیت‌هایی که باید انجام شود، به همراه شرح جزئیات انجام کار و همچنین، ابزاری یکپارچه برای حمایت از اجرای فعالیت‌ها ارائه کرده است. ارتباط بین فعالیت‌های شناسایی شده در چارچوب مقایسه پیشنهادی (جدول ۱) و فعالیت‌های شناسایی شده «انتیرانداک» به‌صورت خلاصه به شرح زیر است:

- ◇ فعالیت‌های مدیریتی: فعالیت مدیریت پروژه؛
- ◇ فعالیت‌های پیش از توسعه: شناسایی منابع دانش؛
- ◇ فعالیت‌های توسعه: فعالیت‌های توصیف نیازمندی‌ها، استخراج دانش خبرگان، استخراج واژگان، مهندسی معکوس، ادغام هستان‌نگارها، مفهوم‌سازی؛
- ◇ فعالیت‌های پس از توسعه: ارتقا و نگهداری؛
- ◇ فعالیت‌های پشتیبانی: شناسایی خبرگان، شناسایی هستان‌نگارهای مرتبط، شناسایی پایگاه داده‌های مرتبط، شناسایی مستندات غیرساخت یافته، ارزیابی خبرگان، مدیریت پروژه، ادغام هستان‌نگارها.

در خصوص معیار حمایت از زبان دیگر، «انتیرانداک» با برچسب‌گذاری فارسی عناصر هستان‌نگار، امکان تحلیل‌های معنایی و ساختاری برای ادغام عناصر هستان‌نگاری را فراهم می‌کند. همچنین، در استخراج واژگان، ابزاری برای پالایش و ایجاد سازگاری در فهرست واژگان بر اساس مشابهت‌یابی ساختاری و معنایی فارسی دارد. مزیت قابل توجه دیگر «انتیرانداک»، تولید فرم‌های ارزیابی از خبرگان بر اساس گزاره‌های فارسی مبتنی بر برچسب‌های فارسی عناصر است.

در نهایت، روش «انتیرانداک» بر اساس تجربیات عملی ساخت هستان‌نگار، جزئیات کافی برای انجام فعالیت‌ها و چگونگی عملکرد پیمان‌های نرم‌افزاری را دارد و ابزار جامعی نیز برای انجام فرایند ساخت ارائه می‌کند.

«انتیرانداک» به‌عنوان یک روش جدید توسعه هستان‌نگار با بهره‌گیری از ایده‌های روش‌های موجود و افزودن ویژگی‌های جدید (مانند پشتیبانی از زبان فارسی، پیمان‌های بودن معماری، توجه به توسعه مشارکتی و یکپارچگی داده‌ها)، طراحی شده و بر اساس

تجربیات حاصل از اجرای دو پروژه ساخت هستان نگار بومی توسعه یافته و در عمل کارایی خود را نشان داده است. پیمان‌های نرم‌افزاری «انتیرانداک» با استفاده از زبان php و پایگاه داده MySQL به صورت تحت وب پیاده‌سازی شده است. این سامانه به صورت متن باز است و طراحی پیمان‌های آن این قابلیت را فراهم آورده است که سایر پژوهشگران و علاقه‌مندان، بخش‌های مورد نظر خود را در آن اضافه کرده و الگوریتم‌ها و ویژگی‌های آن را شخصی‌سازی کنند.

فهرست منابع

- Blázquez, J., M. Fernández, J. M. García-Pinar, and A. Gómez-Pérez. 1998. Building ontologies at the knowledge level using the ontology design environment. *Proceedings of the 11th Banff Knowledge Acquisition for Knowledge-Based Systems*. Banff, Alberta, Canada (pp. 4.1-4.15)
- Brost, W. N. 1997. *Construction of engineering ontologies*. Enschede: University of Twente, Center for Telematica and Information Technology.
- Corcho, O., M. Fernández-López, A. Gómez-Pérez and A. López-Cima. 2005. *Building legal ontologies with METHONTOLOGY and WebODE: 142-157*. Berlin, Heidelberg: Springer.
- De Nicola, A., M. Missikoff and R. Navigli. 2009. A software engineering approach to ontology building. *Information systems* 34 (2): 258-275.
- Fernández, M, A Gómez-Pérez and N. Juristo. 1997. *METHONTOLOGY: From Ontological Art towards Ontological Engineering*. AAAI Spring Symp. Series. Menlo Park, Calif: AAAI Press: 33-40.
- Fernández-López, M. And A. Gómez-Pérez. 2002. Overview and analysis of methodologies for building ontologies. *The Knowledge Engineering Review* 17 (02): 129-156.
- Gailly, F .and G. Poels. 2010. Conceptual modeling using domain ontologies: Improving the domain-specific quality of conceptual schemas. *Presented at the 10th Workshop on Domain-Specific Modeling*. DSM '10. New York, USA: ACM.
- Gómez-Pérez, A., M. Fernández-López and O. Corcho. 2004. Methodologies and methods for building ontologies. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web: 107-197*. Berlin, Heidelberg: Springer.
- Gruber, T. R. 1993. A translation approach to portable ontology specifications. *Knowledge acquisition* 5.199-220 :(2)
- 1995 ._____. Toward principles for the design of ontologies used for knowledge sharing. *International journal of human-computer studies* 43 (5): 907-928.
- Grüniger, M. And M. S. Fox. 1995. Methodology for the Design and Evaluation of Ontologies. *Presented at the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95*. Montreal.
- Guarino, N., D. Oberle and S. Staab. 2009. What is an Ontology? In *Handbook on ontologies*. Springer: Berlin Heidelberg. 1-17.
- Lenat, D. B., R. V. Guha, K. Pittman, D. Pratt and M. Shepherd. 1990. Cyc: toward programs with common sense. *Communications of the ACM* 33 (8): 30-49.
- Noy, N. F and D. L. McGuinness. 2001. *Ontology development 101: A guide to creating your first ontology*. Palo Alto: Stanford Medical Informatics.
- Noy, N. F and M. A. Musen. 2003. The PROMPT suite: interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies* 59 (6): 983-1024.

- Noy, N. F., R. W. Ferguson and M. A. Musen. 2000. The knowledge model of Protege-2000: Combining interoperability and flexibility. In *International Conference on Knowledge Engineering and Knowledge Management*. Berlin Heidelberg: Springer. (pp 17-32).
- Paolucci, M., T. Kawamura, T. R. Payne and K. Sycara. 2002. Semantic matching of web services capabilities. In *The Semantic Web—ISWC 2002*. Berlin Heidelberg: Springer. (pp 333-347).
- Pinto, H. S., C. Tempich and S. Staab. 2009. Ontology engineering and evolution in a distributed world using DILIGENT. *Handbook on ontologies*: 153-176. Berlin Heidelberg: Springer.
- Studer, R., V. R. Benjamins and D. Fensel. 1998. Knowledge engineering: principles and methods. *Data & knowledge engineering* 25 (1): 161-197.
- Suarez-Figueroa, M. C. 2010. NeOn methodology for building ontology networks: Specification, scheduling and reuse (Doctoral dissertation). Madrid: Univ. Politécnica.
- Sure, Y. and R. Studer. 2002. On-To-Knowledge methodology. Technical Report. Karlsruhe, Germany: Institute AIFB, University of Karlsruhe.
- Sure, Y., M. Erdmann, J. Angele, S. Staab, R. Studer and D. Wenke. 2002. OntoEdit: Collaborative ontology development for the semantic web. In *International Semantic Web Conference*. Berlin Heidelberg: Springer. (pp 221-235).
- Uschold, M. and M. Gruninger. 1996. Ontologies: Principles, methods and applications. *The knowledge Engineering Review* 11 (02): 93-136.
- Uschold, M. and M. King. 1995. towards a methodology for building ontologies. *Workshop Held in Conjunction with IJCAI on Basic Ontological Issues in Knowledge Sharing*, Montreal.
- Zaremba, I. 2015. Automatic Transformation of Relational Database Schema into OWL Ontologies. In *Environment. Technology. Resources. Proceedings of the International Scientific and Practical Conference* (Vol. 3) (pp 217-222), Rezekne, Latvia.

امید میلانی فرد

متولد سال ۱۳۵۶، دانشجو دکتری مدیریت فناوری اطلاعات است. ایشان هم‌اکنون کارشناس ارشد فناوری اطلاعات «دانشگاه فردوسی مشهد» است.
سامانه اطلاعات مدیریت، انبار داده، مدل‌سازی و مهندسی داده از جمله علایق پژوهشی وی است.



محسن کاهانی

متولد سال ۱۳۴۶، دارای مدرک دکتری مهندسی کامپیوتر از «دانشگاه ولونگونگ استرالیا» است. ایشان هم‌اکنون استاد تمام گروه مهندسی کامپیوتر دانشگاه فردوسی مشهد و مدیر آزمایشگاه‌های تخصصی آ‌پا و فناوری وب است.
وب معنایی و فرایند‌کاوی از جمله علایق پژوهشی وی است.

