به‌سوی طراحی شبکهٔ واژگانیِ صفاتِ زبان فارسی

مصطفی عاصی[1]
دانشیار پژوهشگاه علوم انسانی و مطالعات فرهنگی
علی رضاقلی‌فامیان[2]
دانشجوی دکتری پژوهشگاه علوم انسانی و مطالعات فرهنگی
داریوش آقاجانی[3]

چکیده

ظهور و گسترش ابزارهای الکترونیکی مربوط به گردآوری، ذخیره‌سازی و پردازش داده‌های زبانی و همچنین موفقیت چشم‌گیر شبکه‌های واژگانی، نگارندگان را بر آن داشت که به طراحی شبکهٔ واژگانی صفات زبان فارسی اقدام کنند. مقالهٔ حاضر ضمن بیان تاریخچه و مبانی کلی شبکهٔ واژگانی، چگونگی شکل‌گیری شبکهٔ حاضر را شرح می‌دهد. از آنجا که قصد طراحی شبکهٔ واژگانی صفات فارسی را داریم، به طبقه‌بندی معنایی مقولهٔ صفت می‌پردازیم. در این طبقه‌بندی ۱۵ طبقهٔ اصلی و بیش از ۷۰ طبقهٔ فرعی شناسایی شده است. برای استخراج صفات و تعیین حوزه‌های معنایی و نیز هم‌معناها، متقابل‌ها و بسامد آن‌ها از سه فرهنگ واژگانی و یک دادگان زبان فارسی استفاده شده است. شبکهٔ واژگانی، نوعی پایگاه داده‌هاست و از این‌رو ضمن آشنایی مقدماتی با مفاهیم بنیادی و روش‌شناسی دادگان با فرآیند طراحی منطقی و فیزیکی شبکهٔ واژگانی صفات فارسی آشنا می‌شویم.

واژه‌های کلیدی: شبکهٔ واژگانی، صفت، هم‌معنایی، تقابل معنایی.

۱- مقدمه

یکی از مهم‌ترین دغدغه‌های انسان در عصر فن‌آوری اطلاعات، طراحی و گسترش ابزارها، امکانات و خدمات مربوط به گردآوری، ذخیره‌سازی و پردازش داده‌های زبانی است. یکی از این امکانات، دادگان زبان طبیعی است که در آن انواع متون یک زبان معیّن در قالب‌های ویژه کدگذاری می‌شود تا بعدها کاربر بتواند بر حسب نیاز کاربردی یا پژوهشی خود اطلاعات مورد نیاز خود را در آن جست و جو کند. شبکهٔ واژگانی، نوعی دادگان واژگانی است و نگارندگان این مقاله می‌کوشند چگونگی طراحی شبکهٔ واژگانی صفات زبان فارسی را شرح دهند. به این منظور ابتدا با ظهور و گسترش شبکه‌های واژگانی در جهان آشنا می‌شویم. در بخش صفت و طبقه‌بندی معنایی آن با چارچوبی آشنا می‌شویم که به کمک آن می‌توان صفات را به حدود هفتاد طبقه تقسیم کرد. در بخش سوم، منابع مورد استفاده برای استخراج داده‌های واژگانی معرفی می‌شوند. در بخش چهارم با اصطلاحات و مفاهیمی مانند دادگان، جدول، رابطه، یکسان‌سازی و معماری دادگان آشنا می‌شویم و در بخش پایانی نیز نحوهٔ درج داده‌ها در نرم‌افزار شبکهٔ واژگانی عرضه می‌شود.

[1] s_m_assi@ihcs.ac.ir
[2] famianali@yahoo.com
[3] daruosha@daruosha.com

۲- تاریخچهٔ شبکهٔ واژگانی

شبکهٔ واژگانی، نوعی دادگان است که با الهام از نظریه‌های روان‌شناسیِ زبان در زمینهٔ حافظهٔ واژگانی انسان سامان یافته است. در شبکهٔ واژگانی، عناصر واژگانی در گروه‌های هم‌معنا طبقه‌بندی و این گروه‌های هم‌معنا به واسطهٔ یک رشته روابط معنایی به یکدیگر متصل می‌شوند. به این ترتیب شبکهٔ واژگانی، شبکه‌ای از روابط معنایی است. از آنجا که رابطهٔ معنایی، رابطه‌ای میان معانی است و از آنجا که معانی در قالب گروه‌های هم‌معنا متجلی می‌شود، می‌توان چنین فرض کرد که روابط معنایی در حکم اشاره‌گرهای بین دسته‌های هم‌معنا هستند. ویژگی دیگر شبکهٔ واژگانی این است که عناصر واژگانی نه بر اساس معنای واژه، بلکه بر پایهٔ صورت واژه سامان‌دهی می‌شوند؛ به همین علت، شبکهٔ واژگانی بیشتر شبیه تزاروس است تا فرهنگ واژه. در مراحل عملیِ طراحیِ شبکه‌های واژگانی نیز، محققان بیشتر از تزاروس‌ها سود برده‌اند.

شبکهٔ واژگانی زبان انگلیسی برای نخستین‌بار در آزمایشگاه علومِ شناختیِ دانشگاه پرینستون، زیر نظر *جرج‌ای. میلر*[1] طراحی شد. این طرح در سال ۱۹۸۵ آغاز شد و برخی نهادهای دولتی دست‌اندرکار طرح‌های ترجمهٔ ماشینی از آن حمایت کردند. انجمن جهانی شبکهٔ واژگانی، نهادی غیرتجاری است که ضمن فراهم کردن شرایط مناسب برای تبادل افکار، اطلاع‌رسانی در خصوص شبکهٔ واژگانی، برگزاری کارگاه‌ها و همایش‌های مختلف، به راه‌اندازی طرح‌های مختلف این حوزه به‌خصوص طراحی شبکه‌های واژگانی زبان‌های مختلف کمک می‌کند. انجمن مذکور دارای وب‌گاه اینترنتی ویژه‌ای است که دانشگاه پرینستون آن را اداره می‌کند و به تمام کاربران اجازه می‌دهد نسخه‌ای از این برنامه را به صورت مستقیم دریافت کنند[2].

آشکارترین تفاوت شبکهٔ واژگانی با فرهنگ‌های زبانی رایج، این است که در شبکهٔ واژگانی، واژگان به چهار مقولهٔ اسم، فعل، صفت و قید تقسیم می‌شوند و عملاً از مقولات نقشی مانند ضمایر خبری نیست. در این شبکه که بازنمودی از حافظهٔ واژگانی ذهن انسان به شمار می‌آید، اسم‌ها در قالب سلسله مراتب موضوعی، افعال با تکیه بر انواع روابط التزامی و صفات و قیدها نیز بر اساس فضاهای چند بعدی سامان‌دهی می‌شوند.

رشد و گسترش ابزارهای رایانه‌ای و خودکار زبانی، از جمله فرهنگ‌های زبانی رایانه‌ای و هم‌چنین موفقیت طرح‌های مربوط به طراحی شبکه‌های واژگانی زبان‌های انگلیسی و زبان‌های عمدهٔ اروپایی، باعث شد این طرح در سایر نقاط جهان نیز با استقبال مواجه شود. به گزارش انجمن جهانی شبکهٔ واژگانی در حال حاضر تلاش برای طراحی یا تکمیل شبکه‌های واژگانی برای زبان‌های مختلف از جمله باسک، بلغاری، عبری، ماراتی، هندی، کاتالا، مولداویایی، روسی، تامیل، ترکی، تایلندی و ویتنامی ادامه دارد. طراحی شبکهٔ واژگانی برای اسم‌ها و افعال زبان فارسی نیز دردانشگاه پرینستون ایالات متحدهٔ امریکا آغاز شده است. محققان طرح مذکور از عنوان و چارچوب پژوهش حاضر مطلعند و در خصوص همکاری دو جانبه اعلام همکاری کرده‌اند.

از آنجا که شبکهٔ واژگانی، نوعی شبکهٔ معنایی محسوب می‌شود و هم‌چنین از آنجا که این نوشتار گزارشی از طراحی شبکهٔ واژگانی صفات زبان فارسی است، ضروری است به معناشناسی مقولهٔ صفت در زبان فارسی بپردازیم.

---

[1] George A. Miller
[2] www.globalwordnet.com

۳- مقولهٔ صفت و طبقه‌بندی‌معنایی آن

صفت در کنار اسم، فعل و قید چهار مقولهٔ اصلی واژگانی را تشکیل می‌دهند. تاکنون طبقه‌بندی‌های مختلفی از صفات عرضه شده است؛ مانند طبقه‌بندی *دیکسون*[1] به سال ۱۹۸۲ که صفات را در ده طبقهٔ معنایی جای داده و طبقه‌بندی *لی*[2] در سال ۱۹۹۴ که ۲۴ طبقه را شناسایی کرده است. نکتهٔ قابل توجه در رویکرد *دیکسون* و *لی* این است که طبقه‌بندی آنان اساساً برای مطالعات فرهنگ‌نگاری و یا رایانه‌ای عرضه نشده است. اکنون به معرفی رویکرد *هوند نرشر ـ اسپلت*[3] می‌پردازیم. این طبقه‌بندی با اندکی تغییر برای طراحی شبکهٔ واژگانی صفت در زبان آلمانی مورد استفاده قرار گرفته است و در ادامه، این طبقه‌بندی را با مثال‌های فارسی عرضه می‌کنیم. گفتنی است رویکرد نگارندگان پژوهش حاضر در طبقه‌بندی صفات زبان فارسی نیز مبتنی بر همین رویکرد است.

۳-۱ رویکرد هوند نرشر ـ اسپلت

۱ـ صفات حسی

روشنایی: تاریک

رنگ: آبی

صدا: بلند

مزه: ترش

بویایی: متعفن

لامسه: زبر

۲ـ صفات زمانی

زمان: د یر

شتاب: چابک

عمر: جوان

عادت: معمولی

۳ـ صفات مکانی

بعد: دراز

جهت: چپ

موقعیت مکانی: نزدیک

منشأ: خارجی

توزیع مکانی: خالی

[1] Dixon
[2] Lee
[3] F. Hundsnurscher-Splett

شکل: گرد
حضور: حاضر

۴ـ صفات حرکت
حرکت: ساکن

۵ـ صفات مادی
ترکیب: طلا یی
حالت: مایع
ثبات: ثابت
انسجام: منسجم
رسیدگی: خام
رطوبت: خشک
خلوص: ناب
گرانش: سنگین
فیزیک: مغناطیسی
شیمی: آهنی
حرارت: گرم

۶ـ صفات آب و هوا
آب و هوا: شرجی

۷ـ صفات مربوط به بدن
زندگی: مرده
بنیه: قوی
گرفتاری: بیمار
تمایل / احساس: گرسنه
جنسیت: مرد
ظاهر: زیبا
وضعیت جسمانی: باردار

۸ـ صفات مربوط به روحیه
روحیه: بشاش

۹ـ صفات مربوط به ذهن و روان
هوش / توجه: باهوش
دانش / تجربه: آگاه

۱۰ـ صفات رفتاری
رفتار / شخصیت: تنبل
جانور: رام
شیوه: عجیب
مهارت: ماهر
روابط: دشمن
همدردی: عزیز
تمایل: خوش صحبت

۱۱ـ صفات اجتماعی
قشر اجتماعی: فقیر
نهاد / سیاست: خصوصی
مذهب: مسیحی
نژاد: زرد
سرزمین: ایران
ناحیه: ملی

۱۲ـ صفات کمیت
شما ر: دو
کمیت: کم
نرخ: گران
بازگشت: حاصلخیز

۱۳ـ صفات رابطه‌ای
اعتبار: معتبر
قطعیت: قطعی
ضرورت: مهم
تأثیر: مؤثر
دشواری: سخت

عملکرد: سالم
ایمنی: خطرناک
نظم: منظم
اتصال: آزاد
تناظر: شبیه
دقت: واضح
اتمام: کامل
ارجاع: مستقیم
سودمندی: مفید

۱۴ـ صفات عمومی
مقایسه: بهتر
ارزیابی: خوب
هنجار: غیرعادی

۱۵ـ صفات مربوط به اسم
بیمارستانی

۳-۲ منابع زبانی

برای گردآوری داده‌های مورد نیاز برای درج در شبکهٔ واژگانیِ صفات فارسی، دو شیوهٔ دستی و خودکار مورد نظر است. در شیوهٔ دستی، سه فرهنگ فارسی و در روش خودکار نیز پایگاه داده‌های زبان فارسی، طراحی شده در پژوهشگاه علوم انسانی و مطالعات فرهنگی، مورد استفاده قرار گرفته است. در ادامه، به معرفی فرهنگ‌های فارسی و پایگاه داده‌ها پرداخته، سپس چگونگی استخراج داده‌ها از منابع مذکور را شرح می‌دهیم.

۳-۲-۱ فرهنگ‌های فارسی

مهم‌ترین منابع نگارنده برای شناساییِ صفات و عرضهٔ هم‌معناها و متقابل‌ها، فرهنگ‌های واژگانی موجود بوده‌اند. در این خصوص سه فرهنگ به‌عنوان منابع اصلی انتخاب شدند که در ادامه به معرفی تک‌تک آن‌ها می‌پردازیم.

۳-۲- ۱-۱ فرهنگ سخن

فرهنگ سخن درسال ۱۳۸۲ و در هشت جلد منتشر شد. این اثر با توجه به اصول نظری و هم‌چنین ملاحظات کاربردی فرهنگ‌نگاری، از بهترین آثار موجود فارسی معاصر به شمار می‌آید. منابع گردآوری،

حوزه‌های مورد نظر زبانی، گزینش واژه‌ها و ریشه‌شناسی با دقت انتخاب شده و هویت دستوری و تعریف مدخل‌ها نیز دقیق و حساب شده است. در این فرهنگ، در تعریف مدخل‌های صفت از الگویی منسجم استفاده شده است که در ادامه به آنها می‌پردازیم:

ـ تعریف با صفت‌هایی مانندِ دارای، فاقد، قابل، غیرقابل، موجب و باعث شروع می‌شود. (ارجمند: دارای قدر و منزلت)

ـ با "آن که" یا " آنچه" شروع می‌شود و سپس یک مترادف وصفی می‌آید. (بیمار: آن‌که دچار نارسایی، آسیب یا اختلال جسمی یا روانی شده باشد؛ مریض)

ـ با "ویژگی" شروع می‌شود. (سر به تو: ویژگی آن‌که راز خود را به کسی نمی‌گوید.)

ـ با مترادف وصفی معنا می‌شود. (بهره‌کش: استثمارگر)

ـ تقدم و تأخر معنایی در درجهٔ نخست بر اساس امروزی یا قدیمی بودن واژه و سپس بر اساس بسامد آن بوده است. معانی پر استعمال، نخست و معانی کم استعمال در مرتبهٔ بعد ذکر شده‌اند.

گفتنی است در گردآوری داده‌ها از *فرهنگ سخن*، به نسخهٔ دو جلدی آن مراجعه می‌شود.

۳-۲-۱-۲ فرهنگ فارسی امروز (ویرایش سوم)

فرهنگ فارسی امروز را برای نخستین بار در سال ۱۳۶۹ مؤسسهٔ نشر کلمه منتشرکرد و ویرایش سوم آن در سال ۱۳۷۷ روانهٔ بازار شد. در این فرهنگ عمومی، هر واژه و تفاوتش با مشابه یا مترادف آن با کوتاه‌ترین جمله‌ها معرفی شده است. همچنین اطلاعاتی در خصوص حالت دستوری، چگونگی جمع‌بستن، ترکیب‌ها و تعلق واژه به مقولهٔ گفتاری، ادبی، کنایی، مجازی و غیره نشان داده شده است. واژه‌های علمی، فنی، محلی یا عامیانه که وارد زبان فارسی معیار نشده‌اند، در محدودهٔ این فرهنگ قرار نگرفته‌اند. نام‌های خاص نیز بنابر قاعده، ذکر نشده‌اند. مؤلفان *فرهنگ فارسی امروز* واژه‌هایی را که در گذشته در شعر یا نثر به کار می‌رفت و امروزه کاربردی ندارند، نیز کنار گذاشته‌اند. از دادن ریشه‌های واژه‌ها نیز خودداری شده است.

این فرهنگ شامل ۵۰۰۰۰ واژهٔ رایج و کارآمد فارسی و ۳۲۵۰۰ مدخل اصلی و ارجاعی است که در گردآوری‌شان از این منابع استفاده شده است: فرهنگ معین، فرهنگ عمید، دایرةالمعارف مصاحب، شماره‌های مختلفی از روزنامه‌ها و مجله‌ها، شماره‌های مختلفی از مجله‌ها و نشریه‌های اختصاصی، کتاب‌های درسی از اول ابتدایی تا چهارم دبیرستان و برخی کتاب‌های اختصاصی مانند نجوم، آشپزی، عکاسی و امثال آن.

۳-۲-۱-۳ فرهنگ جامع واژگان مترادف و متضاد زبان فارسی

فرهنگ جامع واژگان مترادف و متضاد زبان فارسی که آن را به اختصار، *فرهنگِ خداپرستی* می‌خوانیم، شامل ۱۵۰۰۰ مدخل و ۲۷۴۰۰ حوزهٔ معنایی است. در گزینش مدخل‌ها، ابتدا ۱۲۰۰ واژه از واژگان پایهٔ زبان فارسی انتخاب شده و تمام کلماتی که سه بار یا بیشتر در یک پیکرهٔ از پیش طراحی شده وجود داشته‌اند به‌عنوان مدخل اصلی انتخاب شده‌اند. در عرضهٔ عناصر مترادف نیز از شیوهٔ الفبایی استفاده شده است. در مورد مدخل‌های چندمعنایی، ملاک‌های تقدّم و تأخر حوزه‌های معنایی متعدد، مقولهٔ دستوری و بسامد

منظور شده است. *خداپرستی* در مقدمهٔ فرهنگِ خود از اصول شش‌گانه‌ای نام می‌برد که در انتخاب مترادف‌ها و متضادها موردنظرش بوده‌اند:

ـ اصل عام و حاکم بر مکانیسم واژه‌های مترادف؛

ـ اصل جای‌گزینی؛

ـ شم زبانی؛

ـ اصل مقابله و مقایسه؛

ـ اصل احراز هم‌معنایی از طریق سنجش تضاد و تقابل؛

ـ بهره‌جویی از شیوهٔ شبکه ـ گره.

خداپرستی در ادامه می‌افزاید که شیوهٔ شبکه ـ گره روش قاطع و حاکم او بوده است (خداپرستی، ۱۳۷۶: بیست و نه الی سی و دو).

۳-۲-۲ پایگاه داده‌های زبان فارسی

از سال ۱۳۷۲ کار ایجاد پایگاه داده‌های زبان فارسی در پژوهشگاه علوم انسانی و مطالعات فرهنگی آغاز شده است. در این طرح، متون مختلف و متعددی از آثار ادبی، علمی، مطبوعاتی و درسی گرفته تا گفتار طبیعی فارسی زبانان گردآوری شده و در قالب‌های گوناگون در پایگاه داده‌ها ذخیره شده است. داده‌های مذکور در یک پایگاه داده‌های پیوندی[1] ذخیره شده‌اند و این پایگاه توان ذخیره‌سازی ۱۰۰ میلیون واژه را دارد. انواع جست و جو در این پایگاه میسر است: جست و جوی واژگانی، مفهومی، تلفظی، هم‌بافت و غیره. این جست و جوها را می‌توان در محدوده‌های دلخواه (مثلاً دورهٔ زمانی معیّن یا نویسنده‌ای مشخص یا حجم معیّنی از پیکره) انجام داد. پیکرهٔ مورد بحث ما انواع امکانات پژوهشی و کاربردی را فراهم می‌کند؛ از جمله امکان تهیهٔ فهرست‌های آماری و بسامدی از واژه‌های متون، امکان اجرای عملیات برچسب‌دهی دستی یا خودکار و امکان عرضهٔ خدمات و اطلاعات و گزارش‌های مختلف به کاربران و پژوهندگان ایرانی و جهانی به صورت برخط و برون خط.

در حال حاضر پایگاه داده‌های زبان فارسی حاوی ۳۶ میلیون واژه است و منبع ارزشمندی برای نگارندگان پژوهش حاضر محسوب می‌شود؛ چراکه به کمک آن می‌توان در خصوص بسامد صفات فارسی با دقت و قاطعیت سخن گفت و هم‌چنین نمونه‌های مناسبی از هر صفت در بافت طبیعی و کاربردی‌اش را در قالب مثال ذکر کرد.

۴- طراحی و اجرای پایگاه داده‌ها

۴-۱ مفاهیم و اصطلاحات بنیادی

۴-۱-۱ پایگاه داده‌ها

الگوی پایگاه داده‌های پیوندی را ابتدا *ای. اف. کاد*[2] مطرح کرد. این الگو مبتنی بر شاخه‌هایی از ریاضیات، موسوم به نظریهٔ مجموعه‌ها و منطق گزاره‌ها طراحی شده است. گفتنی است منظور از "رابطه‌ای" این نیست

[1] Relational Database
[2] E. F. Codd

که جدول‌ها با هم ارتباط دارند؛ البته ارتباط دارند؛ اما منظور این نیست. "رابطه‌ای" اصطلاحی است که *کاد* در توضیح نظریه‌اش از آن استفاده کرد. در نوشته‌های *کاد* منظور از رابطه، جدول است و جدول نیز یعنی مجموعه‌ای مرتبط از اطلاعات.

پایگاه داده‌ها به‌طور خلاصه "مجموعه‌ای از داده‌هاست که به گونه‌ای سامان‌دهی شده‌اند که به سهولت و سرعت بتوان آنها را جستجو و بازیابی کرد". (دیوید سن، ۲۰۰۱: ۴۸)

در تعریفی کامل و جامع می‌توان گفت:

پایگاه داده‌ها مجموعه‌ای است از داده‌های ذخیره شده و پایا، به‌صورت مجتمع (یکپارچه) (نه لزوماً به‌طور فیزیکی، بلکه حداقل به‌طور منطقی)، به هم مرتبط، با کمترین افزونگی[1] ، دارای یک طرح منطقی مبتنی بر یک مدل داده‌ای و توصیف‌شده به‌ویژه در محیطی انتزاعی؛ یعنی در چارچوب همان مدل داده‌ای، تحت مدیریت یک سیستم کنترل متمرکز، مورد استفادهٔ یک یا چند کاربر از یک یا بیش از یک سیستم کاربردی، به‌طور هم‌زمان و اشتراکی. (روحانی رانکوهی: ۱۸).

پایگاه داده‌های رابطه‌ای امتیازات عمده‌ای دارد که به‌طور اجمالی به آن‌ها اشاره می‌کنیم:

ـ درج مدخل، روزآمدکردن و حذف آن از کارایی بالایی برخوردار است؛

ـ بازیابی، خلاصه‌سازی و گزارش‌دهی، ساده است؛

ـ از آنجاکه پایگاه داده‌ها الگویی خوش‌ساخت و قاعده‌مند را دنبال می‌کند، پیش‌بینی‌شونده است؛

ـ اِعمال تغییر در پایگاه داده‌ها آسان است.

۴-۱-۲ جدول

در الگوی پایگاه داده‌ها، جدول برای نمایش هستینه‌ها[2] در جهان خارج به‌کار می‌رود. "هستینه" عبارت است از هرگونه شیء یا رویداد جهان خارج؛ مانند مشتری یک رستوران، سفارش غذا یا واژه.

در الگوی پایگاه داده‌های رابطه‌ای، هر جدول، از ردیف‌ها [3]و ستون‌ها[4] تشکیل می‌شود. هر ردیف از جدول باید یگانه باشد. چنانچه این اصل رعایت نشود، نمی‌توان ازطریق برنامه‌نویسی به یک ردیف معیّن از جدول ارجاع کرد. بدیهی است چنین حالتی منجر به انواع ابهام و مشکلات خواهد شد. یگانگی جدول وقتی محقق می‌شود که به هر جدول، یک کلید اصلی اختصاص دهیم. منظور از کلید اصلی، ستونی است که برای جدول، ارزش‌های یگانه دارد. هر جدول می‌تواند فقط یک کلید اصلی داشته باشد. به تمام ستون‌ها یا ترکیبی از ستون‌های جدول اصطلاحاً "کلیدهای منتخب" گفته می‌شود. کلید اصلی از میان کلیدهای منتخب برگزیده می‌شود و به سایر کلیدها اصطلاحاً "کلیدهای ثانوی" گفته می‌شود. کلیدها ممکن است ساده یا مرکب باشند. کلید ساده فقط از یک ستون تشکیل شده؛ در حالی‌که کلید مرکب شامل دو یا چند ستون است.

در انتخاب کلید اصلی از میان کلیدهای منتخب باید به این اصول توجه کرد:

ـ کمینه‌گرایی: انتخاب کم‌ترین ستون مورد نیاز؛

---

[1] Redundancy
[2] Entity
[3] Rows
[4] Columns

ـ ثبات: ستون انتخاب شده به ندرت دچار تغییر شود؛
ـ سادگی/ آشنایی: کلیدی انتخاب شود که برای کاربر ساده و آشنا باشد.

در اغلب موارد از عدد صحیح (نه متن توصیفی) استفاده می‌شود؛ چراکه امکان بروز اشتباه در حروف یا تغییر نام وجود ندارد.
اصطلاح دیگرِ مربوط به جدول، عبارت است از "کلید خارجی"[1]. کلید خارجی؛ یعنی ستونی از یک جدول که برای ارجاع به کلید اصلیِ جدولی دیگر به‌کار می‌رود.
نکتهٔ حائز اهمیت این است که کلیدهای خارجی و کلیدهای اصلی باید ارزش‌هایشان را از حوزه‌ای مشترک دریافت کنند. حوزه در اصطلاح، گنجینه‌ای از ارزش‌هاست که ستون‌ها را از آن انتخاب می‌کنیم.

۴-۱-۳ روابط

در یک پایگاه داده‌ها، کلیدهای خارجی را به‌گونه‌ای تعریف می‌کنند که بازنمود روابط در جهان خارج باشند. واقعیت این است که روابط بین هستینه‌های واقعی جهان خارج بسیار پیچیده و چندگانه است؛ اما در یک پایگاه داده‌ها روابط فقط بین دو جدول تعریف می‌شود. بین دو جدول یکی از این سه نوع رابطه برقرار است: یک به یک، یک به چند و چند به چند. در ادامه تعریف *لیتوین* از روابط را عرضه می‌کنیم (لیتوین: ۵-۴).

۴-۱-۳-۱ رابطهٔ یک به یک

اگر برای هر ردیفِ جدول اول، حداکثر یک ردیف در جدول دوم موجود باشد؛ دو جدول دارای رابطهٔ یک به یک هستند؛ البته رابطه واقعی یک به یک به ندرت در جهان خارج مشاهده می‌شود و در پایگاه داده‌ها نیز به ندرت به‌کار می‌رود. پایگاه داده‌های شبکهٔ واژگانی صفات فارسی نیز فاقد چنین رابطه‌ای است.

۴-۱-۳-۲ رابطهٔ یک به چند

دو جدول با یکدیگر رابطهٔ یک به چند دارند، اگر برای هر ردیف در جدولِ اول، چند ردیف در جدول دوم؛ اما برای هر ردیفِ جدول دوم فقط یک ردیف در جدول اول موجود باشد. در پایگاه داده‌های شبکهٔ واژگانی، در بسیاری از موارد، یک واژه با چند هم‌معنا و به بیان فنی‌تر با چند ردیف در جدول هم‌معناها مرتبط می‌شود. صفتِ "سرخ" به‌عنوان یک مدخل با سه معنای "سرخ رنگ"، "خونین" و "کمونیست" مرتبط می‌شود و لذا نمونه‌ای مناسب از رابطهٔ یک به چند محسوب می‌شود.

۴-۱-۳-۳ رابطهٔ چند به چند

دو جدول دارای رابطهٔ چند به چند هستند، اگر برای هر ردیف در جدول اول، ردیف‌های متعددی در جدول دوم و برای هر ردیف جدول دوم ردیف‌های متعددی در جدول اول موجود باشد. در شبکهٔ واژگانیِ صفاتِ فارسی، واژهٔ "روشن" با دو واژهٔ "درخشان" و " فعال" هم‌معناست و از طرف دیگر خود واژهٔ "درخشان" با کلماتی غیر از"روشن" نیز هم‌معناست. این نمونه‌ای از رابطهٔ چند به چند به‌شمار می‌آید.

[1] Foreign Key

۴-۱-۴ یکسان‌سازی

در طراحی پایگاه داده‌ها پرسش‌های مختلفی مطرح می‌شود؛ برای مثال: به چند جدول نیاز داریم و این جدول‌ها بازنمود چه هستینه‌هایی هستند؟ چه ستون‌هایی در جدول‌ها درج خواهد شد؟ رابطهٔ میان جدول‌ها چه خواهد بود؟ پاسخ این پرسش‌ها را باید در مفهومِ "یکسان‌سازی" جست و جو کرد. یکسان‌سازی عبارت است از فرایندِ ساده‌سازیِ طراحی پایگاه داده‌ها به‌طوری‌که به ساختاری بهینه و عاری از افزونگی دست یابیم. در یکسان‌سازی، مفهوم صورت‌های هنجار مطرح است. منظور از صورت‌های هنجار، تسلسلی خطی از قواعد است که در پایگاه داده‌ها به‌کار می‌روند و با هر صورتِ هنجار بالاتر می‌توان به طراحی بهتر و کارآمدتری دست یافت. صورت‌های هنجار بر پایهٔ روابط استوارند و نه جدول‌ها. در پایگاه داده‌ها منظور از رابطه، جدول ویژه‌ای است که دارای این مشخصه‌ها باشد: (لیتوین: ۵)

ـ رابطه‌های یک هستینه را توصیف می‌کنند؛

ـ ردیف‌های تکراری ندارند؛ بنابراین همیشه یک کلید اصلی وجود دارد؛

ـ ستون‌ها نامرتب هستند؛

ـ ردیف‌ها نامرتب هستند.

۴-۲ معماری پایگاه داده‌ها

پس از آشنایی مختصر با مفاهیم ابتدایی پایگاه داده‌ها به اجمال به ساختار کلی پایگاه داده‌های شبکهٔ واژگانی حاضر می‌پردازیم. در این پایگاه داده‌ها شش جدول در نظر گرفته شده است که عبارتند از: الف) جدول مدخل‌ها که در آن هر مدخل در یک رکورد درج می‌شود؛ ب) جدول توصیف‌ها که شامل توصیف و توضیحات هر مدخل است؛ ج) جدول هم‌معناها که در آن توصیف‌های مشابه مدخل‌ها به یکدیگر ارجاع می‌کنند؛ د) جدول متقابل‌ها که کارکردش شبیه جدول هم‌معناهاست؛ ه) جدول برچسب‌ها که طبقهٔ معنایی صفات را نشان می‌دهد؛ و) جدول اسم‌ها که در مواردی که یک صفت، مترادف یا متضاد ندارد، صفت مذکور را به یک اسم مرتبط می‌کند. در این جدول‌ها، داده‌ها به‌صورت الفبای فارسی و لاتین درج می‌شود و کاربر می‌تواند به دو شیوهٔ الفبایی فارسی و لاتین، اطلاعات را بازیابی کند. محیط نرم‌افزاری این سیستم به گونه‌ای طراحی شده است که کاربر می‌تواند محیط مورد دلخواه خود (مانند ویندوز مایکروسافت، لینوکس و غیره) را انتخاب کند.

۵- طراحی شبکهٔ واژگانیِ صفات زبان فارسی

پس از آشنایی با مقولهٔ صفت و طبقات معنایی آن و نیز طراحی منطقی جدول‌های دادگان، اکنون به بخش اصلی کار خود یعنی استخراج داده‌ها و درج آنها در دادگان طراحی شده می‌پردازیم. برای این‌کار، ابتدا به *فرهنگ سخن* و فرهنگ *فارسی امروز* مراجعه کرده، ضمن شناسایی اقلام صفت در خصوص تعداد حوزه‌های معنایی هر صفت و جزئیات آن تصمیم می‌گیریم. سپس با استفاده از همین فرهنگ‌ها و نیز *فرهنگ خداپرستی* به ردیابی و درج هم‌معناها و متقابل‌های هر مدخل می‌پردازیم. گفتنی است در این مرحله به شم زبانی خود و سایر فارسی زبانان نیز تکیه می‌کنیم. به لحاظ بسامد نیز صفات را در شش طبقهٔ بسیار

پربسامد، پربسامد، رایج، کمی رایج، کم‌بسامد و بسیارکم بسامد طبقه‌بندی می‌کنیم. پیکرهٔ زبان فارسیِ پژوهشگاه علوم انسانی و مطالعات فرهنگی، منبع مناسبی است که بسامد هر صفت را در یک متن معیّن با عدد نشان می‌دهد و با انتخاب دامنه‌ای مشخص می‌توان صفات را در یکی از این طبقات شش‌گانه جای داد. شبکهٔ واژگانی صفات زبان فارسی پس از اتمام شامل حدود ۵۰۰۰ صفت خواهد بود و از این شبکه می‌توان در امور نظری و کاربردیِ متفاوتی از جمله مطالعات و پژوهش‌های مختلف زبان‌شناسی، تدوین فرهنگ‌های واژگانی و تزاروس‌ها، ابهام‌زدایی‌معنایی، تهیهٔ بانک‌های واژگانی نرم‌افزارهای ترجمهٔ ماشینی و غیره استفاده کرد.

منابع

ـ انوری، حسن (۱۳۸۲)، *فرهنگ سخن*، تهران: انتشارات سخن.

ـ خداپرستی، فرج‌اله (۱۳۷۶)، *فرهنگ جامع واژگان مترادف و متضاد زبان فارسی*، دانشنامهٔ فارس.

ـ روحانی رانکوهی، محمدتقی (۱۳۸۰)، *مفاهیم بنیادی پایگاه داده‌ها*. تهران: انتشارات جلوه.

ـ صدری افشار، غلامحسین و دیگران (۱۳۷۷)، *فرهنگ فارسی امروز (ویرایش سوم)*، تهران: نشر کلمه.

ـ عاصی، مصطفی (۱۳۸۴)، *"پایگاه داده‌های زبان فارسی در اینترنت (۱)"*، پژوهشگران، پژوهشگاه علوم انسانی و مطالعات فرهنگی، س ۱۳، ش ۲.

- Azarova, Irina and Anna Sinopalnikova (2004) “Adjectives in RussNet” In: *Proceedings of the Second Global WordNet Conference*, pp 251-258,Brno, Czech Republic.
- Davidson,Louis (2001) *2Professional SQL Server 2000 Database Design*, Wrox Press Ltd.
- Dixon, R. W. (1982) *Where have all the adjectives gone?*, Mouton Publishers,
- Fellbaum,C , et al, (1990) *Adjectives in WordNet*, in G. Miller (ed) “Five papers on WordNet”, *International Journal of Lexicography* 3 (4), 1990
- Hundsnurscher, F. & J. (1982) Splett: *Semantik der Adjektive im Deutschen:Analyse der semantischen Relationen*. Westdeutscher Verlag,
- Lee, S. (1994) *Untersuchungen zur Valenz des Adjektivs in der deutschen Gegenwartssprache*. Berlin: Lang,
- Litwin,Paul, *Fundamentals of Relational Database Design*, Microsoft TechNet