

اطلاع رسانی

نشریه فنی مرکز اطلاعات و مدارک علمی ایران

دوره هفتم؛ شماره ۱

طرح ابتدائی منطق بول^۱ و کاربرد آن در نظامهای بازیابی پس همارا^۲

نوشته: ضیاء موحد
درآمد

در نوشه های محققان دانش اطلاع رسانی، بخصوص در بحث از نظامهای بازیابی پس همارا اشاره هایی به منطق بول می شود. منطق بول، که از نوآوریهای جورج بول^۳ دانشمند انگلیسی است، ساختاری ریاضی است که به اعتبار ساخت صوری آن تعبیری گوناگون می پذیرد. جای بحث منطق بول در منطق ریاضی و ریاضیات جدید است. از این روست که محققان دکومانتاسیون در نوشه های خود اغلب با اشاره کوتاهی به این منطق از آن می گذرند و خواننده برای آگاهیهای بیشتر خود باید به کتابهای مناسب مراجعه کند.

هدف ما در این مقاله طرح ریاضی منطق بول نیست. می خواهیم مفهومهایی از این نظریه را که دانستن آن برای علاقمندان به دانش اطلاع رسانی لازم می نماید با بیانی ساده شرح دهیم و آنگاه به کاربرد آن در یکی از نظامهای ساده بازیابی بپردازیم. این نوشه، زمینه را برای طرح دقیق منطق بول و نظریه های تازه تری که در نظامهای بازیابی کاربرد پیدا کرده اند و فهم آنها به دانستن مبانی نظری بالاتری دارد، فراهم می آورد. یکی از این نظریه ها، نظریه مجموعه هایی است که مرز روشنی^۴ ندارند. این نظریه اثر یکی از محققان ایرانی، پروفسور لطفی زاده است که بخصوص در رفع پاره ای از دشواریهای نظامهای بازیابی که مبتنی بر منطق بول هستند، اهمیت یافته است. درباره این نظریه در شماره های آینده سخن خواهیم گفت.

مجموعه

مجموعه، مفهومی چنان بدیهی و بنیادی می نماید که نمی توان آن را با مفهومی ساده تر از آن تعریف کرد. این که بگوییم:
به هر دسته از شیئها یا به هر گروه از چند شیء مجموعه می گویند.

¹ Boolean Logic

² Post-coordinate Retrieval Systems

³ George Boole

⁴ Fuzzy Sets

تعریفی از مجموعه نکرده ایم. "دسته"، "گروه"، و "چند" خود واژه های دیگری برای مجموعه هستند و از این رو برای فهم مفهومی که این واژه ها بیان می کنند، باید از پیش مفهوم مجموعه را دانسته باشیم. این گونه مفهومها را که از تعریف می گریزند تعریف ناپذیر^۱ می نامند. مجموعه یکی از تعریف ناپذیری های نظریه مجموعه هاست.

مجموعه ای از چند شیء را با قراردادن نام آن چند شیء در داخل ابرو نشان می دهیم. برای مثال:

{حسن، سیب، ایران، ۲}

مجموعه ای است با چهار عضو:^۲ حسن، سیب، ایران و ۲. مجموعه هایی که شماره عضوهای آنها بینهایت یا نامعلوم باشد همیشه نمی توان آنها را به روش بالا نشان داد. عضوهای چنین مجموعه هایی اغلب در صفتی یا صفتی هایی همانندند. برای مثال :

{۰۰۰۰۰۰۰، ۸، ۶، ۴، ۲}

مجموعه عددهایی است که همه زوج هستند. از این رو، این مجموعه را با مفهوم "عدد زوج" می نماییم و مجموعه عددهای زوج را بنا به قرارداد با یک منحنی بسته بیشتر دایره نمایش می دهیم:



هر نقطه ای از این دایره نماینده عددی است زوج و هر عدد زوج عضوی است از این مجموعه. در بنیاد، هر واژه ای که مفهومی را برساند- مفهوم- واژه^۳- مجموعه ای را می نمایاند که عضوهای آن در آن مفهوم همانندند. برای مثال هر کدام از مفهوم- واژه های انسان، حیوان، زن دار، تهران، ایران و ...

نماینده مجموعه ای هستند. این مجموعه را مصادقهای یا دایره مصاديق آن مفهوم هم می نامند. کار ما از این پس بررسی مفهومها و بستگیهای آنها با یکدیگر است. اما پیش از این کار، تذکر چند نکته بی فایده نخواهد بود.

۱) هر مفهوم واژه معنایی^۴ دارد و مصادقی^۵ (یا مصادقهایی) برای مثال می دانیم که سه نیمساز هر مثلث در یک نقطه همدیگر را قطع می کنند، هم چنانکه سه میانه هر مثلث و نقطه اول در هر مثلث همان نقطه دوم است. بنابر این عبارتهای

نقطه برخورد سه نیمساز هر مثلث

¹ Indefinable

² Member

³ Concept -word

⁴ Intension

⁵ Extension

نقطه برخورد سه میانه هر مثلث

از نقطه های یکسانی سخن می گویند. این نقطه ها مجموعه معینی از نقطه های همه مثلثهاست. در واقع مفهومی که با عبارت اول بیان می شود، همان مصادقه را دارد که مفهومی که با عبارت دوم. یعنی این دو مفهوم هم مصدق هستند. اما هر فارسی زبانی می پذیرد که معنای اول یا عبارت دوم یکی نیست. اینجا با دو مفهوم سر و کار داریم که هم مصدق هستند اما هم معنا نیستند. اگر مفهوم چیزی جز مصدق نبود این تفاوت هم وجود نداشت. از این رو معنای مفهوم^۱ را از مصدق مفهوم^۲ جدا می کنیم.

در نظریه مجموعه ها با مصادقه های مفهومها کار داریم نه معنای آنها.

(۲) این که گفتیم هر مفهوم مجموعه ای را می نمایاند حرف دقیقی نیست. مجموعه هایی هستند که هیچ مفهوم_ واژه ای برای آنها در زبان نیست و مفهومهایی هستند که هیچ مجموعه ای را نمایش نمی دهند. بحث در این موردها از سطح این نوشته فراتر می رود.

(۳) بعضی مفهومها مانند "انسان" ، "طلاء" و "آب" مصادقه های روشی دارند. اما بعضی دیگر- در واقع بیشتر مفهومها - مانند "بلند" ، "بازی" ، "با هوش" برخلاف آنچه اغلب گمان می کنند مصادقه های روش و دقیقی ندارند، یعنی دایره مصادقه های آنها مرز قاطعی ندارد. این امر بویژه در نظامهای بازیابی اطلاعات که در آنها مدرکها را با رشته ای از مفهومها می شناسند و بازیابی می کنند پی آمدهای ناخواسته ای دارد. نظریه مجموعه هایی که مرزهای دقیقی ندارند بخصوص برای رفع این دشواری طرح شده است. در این رشته فرض ما این است که دایره مصادقه های هر مفهوم مرز قاطعی دارد. از این پس بیشتر از مجموعه ها (مصادقه ها) سخن خواهیم گفت تا مفهوم ها. خواننده می تواند هر مجموعه را دایره مصادقه های مفهومی تصور کند.

متتم مجموعه

هر مفهوم شیئها را به دو گروه می کند. گروهی که در دایره مصادقه های آن هستند و گروهی که بیرون از این دایره قرار می گیرند. برای مثال مفهوم "انسان" شیئها را به دو گروه انسان و غیر انسان دسته بندی می کند. گروه اول را مجموعه وابسته به مفهوم و گروه دوم را متتم این مجموعه می نامیم. بنابراین متتم هر مجموعه مجموعه ای است که در آن تنها آن شیئه ای باشد که در مجموعه اول نیستند.

با این تعریف متتم هر مجموعه، مجموعه ای با بینهایت عضو خواهد شد برای مثال متتم مجموعه کارمندان نه تنها شامل انسانهای غیر کارمند بلکه شامل تمام شیئه های جهان از ستارگان گرفته تا اتمها خواهد بود. در صورتی که وقتی از کارمندان سخن می گوییم می خواهیم تنها انسانهای کارمند را از غیر کارمند جدا کنیم. از این رو، در این مثال بهتر است نخست مجموعه چیزهایی که می خواهیم درباره

¹ Concept in intension

² Concept in Extension

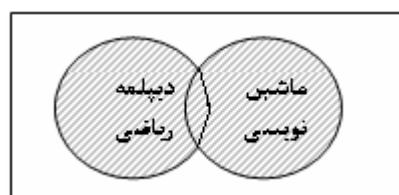
آنها سخن بگوییم برگزینیم، آنگاه متمم مجموعه های کوچکتر در درون این مجموعه را نسبت به آن پیدا کنیم. مجموعه ای که نسبت به آن مجموعه های متمم را پیدا می کنیم مجموعه کلی^۱ می نامیم و به U نشان می دهیم. مجموعه کلی را بیشتر با مستطیل و مجموعه های دیگر را با دایره هایی در آن نشان می دهند.



در نمودار بالا مستطیل، مجموعه انسانها و دایره، مجموعه کارمندان و بخش هاشور خورده، انسانهای غیر کارمند- متمم مجموعه کارمندان - را نشان می دهد، اگر مجموعه کارمندان را با A نشان دهیم متمم آن را بنا به قرارداد با \bar{A} مشخص می کنیم.

جمع دو مجموعه

منظور از جمع دو مجموعه A و B مجموعه ای است که عضوهای آن یا عضو A یا عضو B و یا عضو هر دو باشد. مثال: سازمانی آگهی کرده است که به چند ماشین نویس یا دارنده دیپلم ریاضی نیاز دارد. واضح است که چنین سازمانی افرادی را هم که هم ماشین نویس باشند و هم دارنده دیپلم ریاضی استخدام خواهد کرد. نمودار زیر این را بروشنی نشان می دهد.



در اینجا هم مستطیل همه افراد انسانی و بخش هاشور خورده مجموعه ماشین نویسانی با دارندگان دیپلم ریاضی و بخش خارج از دو دایره متمم این مجموعه را نشان می دهد.
جمع دو مجموعه A و B را با $A \cup B$ نشان می دهیم و آن را A یا B می خوانیم. بسادگی می توان دریافت که جمع هر مجموعه با متمم آن برابر با مجموعه کلی است.

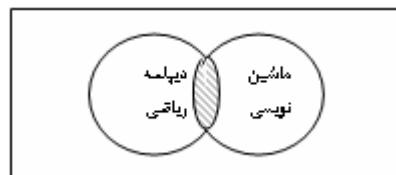
$$A \cup \bar{A} = U$$

ضرب دو مجموعه

منظور از ضرب دو مجموعه A و B مجموعه ای است که عضوهای آن هم عضو A و هم عضو B باشند.

¹ Universal set

مثال: اگر سازمان مذکور آگهی کرده باشد که به افراد ماشین نویس دارنده دیپلم ریاضی نیاز دارد، نه دارندگان دیپلم ریاضی را که ماشین نویسی ندانند خواهد پذیرفت و نه ماشین نویسانی را که دیپلم ریاضی ندارند. نمودار این را نشان می دهد.



در نمودار بالا تنها بخش هاشور خورده نماینده ماشین نویسان دارنده دیپلم ریاضی است. ضرب دو مجموعه A و B را با $A \cap B$ نشان می دهیم و آن را A و B می خوانیم.

مجموعه تهی^۱

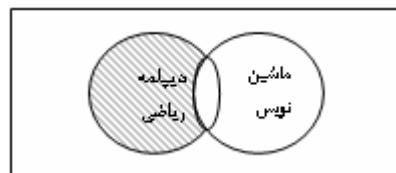
اگر در مثال بالا ماشین نویس دارنده دیپلم ریاضی نداشته باشیم دو دایره A و B هیچ نقطه مشترکی نخواهند داشت. عبارت ریاضی تر می گوییم دو مجموعه A و B در هیچ مشترکند. هیچ را در نظریه مجموعه ها، مجموعه تهی می نامیم، یعنی مجموعه ای که هیچ عضوی ندارد. برای مثال مفهومهای دایره مربع، "انسان سه پا"، "سیمرغ" و "قاف" همه مفهومهایی هستند بی هیچ مصدق. به بیان دیگر دایرة فرضی مصادقهای آنها خالی است. مفهومهایی که مصادقی ندارند مجموعه مصادقهای آنها مجموعه تهی است. مجموعه تهی را با \emptyset نشان می دهیم. اهمیت مجموعه تهی در نظریه مجموعه ها همانند اهمیت صفر در حساب است. درستی رابطه های زیر را با اندک تأملی می توان دریافت.

$$\emptyset = U \quad \bar{U} = \emptyset$$

یعنی متمم مجموعه تهی، مجموعه کلی و متمم مجموعه کلی، مجموعه تهی است.

تفریق دو مجموعه

اگر بخواهیم در مثال بالا مجموعه کسانی را پیدا کنیم که دیپلم ریاضی دارند اما ماشین نویسی نمی دانند باید از مجموعه دارندگان دیپلم ریاضی بخش مشترکی را که با مجموعه ماشین نویسان دارد حذف کنیم.



¹ Null set

بخشها شور خورده این مجموعه را نشان می دهد. اگر A مجموعه دارندگان دیپلم ریاضی و B مجموعه ماشین نویسان باشد، افرادی که عضو A هستند اما عضو B نیستند مجموعه ای را می سازند که به آن تفرقی A و B می گوییم و آن را چنین می نویسیم:

A-B

آنچه را درباره جمع و ضرب و تفرقی دو مجموعه گفتیم بسادگی می توان به بیش از دو مجموعه تعمیم داد. این را با مثالهایی در مورد نظامهای بازیابی پس همارا روشن کنیم. فرض ما این است که خوانندگان با اصول نمایه سازی و روشهای بازیابی همارا آشنا هستند. با این همه اشاره ای کوتاه به روشهای بازیابی همارا بی فایده نخواهد بود.

مرکزهای اسناد و مدارک^۱ گذشته از گردآوری، فهرست برداری و نگاهداری مدرکها نه تنها باید بتوانند هنگام نیاز، هر مدرک را از میان انبوه مدرکهای دیگر براحتی باز بیابند بلکه باید بتوانند (و این از مهمترین وظیفه های مرکز های اسناد است) مدرکهایی را که در موضوعی خاص دارند دقیق و سریع بازیابی کنند. برای مثال اگر مرکز اسنادی مدرکهای مربوط به تاریخ و فرهنگ ایران را گردآوری می کند و پژوهنده ای نیاز به مدرکهایی درباره تاریخ مطبوعات آذربایجان داشته باشد، این مرکز باید بتواند نیاز پژوهنده را با دقت و سرعت برآورد. البته چنین کاری همیشه با کم و کاستیهایی همراه خواهد بود. به این معنی که همراه مدرکهای مناسب و مربوط، همیشه مدرکهای دیگری هم که با موضوع ربط چندانی ندارند بازیابی می شوند. همچنین بعضی از مدرکهای مناسب دور از دسترس می مانند. برای غلبه بر این دشواریهایست که روشهای بازیابی همارا را پیدید آورده اند این روشها به دو گروهند: پیش همارا^۲ و پس همارا. در روش پیش همارا را محتوای هر مدرک را به نوعی با ترکیب چند واژه با هم (که ترتیب پشت سر هم نهادن آنها موضوع اصلی بحث در این روش است) نشان می دهند. برای مثال به هر مدرکی که در آن از تاریخ مطبوعات آذربایجان سخن رفته باشد (اگر بخواهیم مطلب را خیلی ساده کنیم) عنوان ترکیبی "تاریخ مطبوعات آذربایجان" را می دهند و فهرست همه آن مدرکها را به ترتیبی زیر این عنوان می آورند.

روش دیگری روش بازیابی پس هماراست... در یکی از این روشها به جای این که مانند روش بالا عنوان کاملی به مدرک بدھند و مشخصات مدرک را زیر آن بنویسند، تک تک واژه هایی که ترکیب آنها عنوان کامل مدرک را می سازد می یابند و هر واژه را جداگانه در برگه ایی می نویسند. بدین ترتیب مشخصات هر نوشته ای درباره تاریخ مطبوعات آذربایجان" زیر سه واژه "تاریخ"، "مطبوعات" و "آذربایجان" نوشته می شود. در این روش، بر خلاف روش بالا، تنها هنگام بازیابی مدرکهایست که

¹ Documentation Centres

² Pre-coordinate

واژه‌ها را ، به شرحی که هم اکنون خواهیم دید، بر اساس منطق بول با هم ترکیب و مدرکها را بازیابی می‌کنند.

روشهای بازیابی پس همارا به سه گروه تقسیم می‌شوند: دستی، نیمه خودکار و خودکار در اینجا روش بازیابی پس هما را با برگه‌های جدول را که روشنی دستی است و در بسیاری از مراکز اسناد ایران می‌تواند با سهولت و کارآیی بکار رود، کوتاه شرح دهیم.

نظام بازیابی پس همارا با برگه‌های جدول

مرحله‌های اساسی ذخیره و بازیابی مدرکها را در این روش می‌توان چنین خلاصه کرد:

- ۱) مدرکها را در مرکز به ترتیب دریافت (با ترتیب مناسب دیگر) شماره گذاری می‌کنند.
- ۲) متخصصان موضوعی مدرکها را بررسی می‌کنند و واژه‌هایی را که نشان دهنده محتوای هر مدرک است بر کاغذهای از پیش آماده شده ای (کاربرگه) می‌نویسند
- ۳) به هر واژه که در کاربرگه آمده برگه ای به نام برگه جدول که ده ستون دارد و در بالای آن جایی برای نوشتمن واژه نهاده اند، اختصاص می‌دهند و آن واژه را در آنجا می‌نویسند.

برای مثال برگه جدول واژه "تاریخ" می‌تواند چیزی مانند شکل پائین باشد.

تاریخ									
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹
	۱۱	۱۲	۲۳	۹۴	۱۵	۴۵۶	۵۰	۷۸	

چنان که در این شکل دیده می‌شود شماره مدرکهای مناسب را در ستونهایی می‌نویسند که رقم بالای آن ستونها همان آخرین رقم سمت راست شماره مدرک باشد.

- ۴) این برگه‌ها را به ترتیب الفبایی واژه‌های بالا پشت هم می‌گذارند، در اینجا مرحله ذخیره اطلاعات مدرکها به پایان می‌رسد.

در مثال بالا برای نوشته‌ای که درباره تاریخ مطبوعات آذربایجان بود سه برگه جدول تهیه می‌کنند و شماره این مدرک را در هر سه برگه در ستون مناسب می‌نویسند. واضح است که در هر برگه جدول،

شماره های متعدد دیگری نیز نوشته می شود. این شماره ها مربوط به مدرکهای دیگری هستند که محتوای آنها نیز با واژه بالای برگه بستگی پیدا می کند. بنابراین: هر برگه جدول مجموعه ای است که عضوهای آن شماره مدرکهایی هستند که با واژه بالای برگه به نوعی ارتباط معنایی دارند. اکنون به کمک این برگه ها (مجموعه ها) می توانیم انواع ترکیب مجموعه ها را پیدا کنیم. در مثالهای پائین فرض می کنیم که مركز اسنادی داریم که تنها مدرکهای مربوط به ایران را گردآوری می کند یعنی مجموعه کلی ما ، U ، مجموعه مدرکهای مربوط به ایران است. همچنین فرض می کنیم مجموعه شماره هایی که در برگه های جدول مثال ما نوشته شده بشرح زیر باشند:

$$\{11, 2, 23, 12, 94, 456\} = \text{تاریخ}$$

$$\{3, 23, 15, 456, 78, 89\} = \text{مطبوعات}$$

$$\{11, 14, 456, 78, 89\} = \text{آذربایجان}$$

ضرب منطقی

می خواهیم مدرکهایی که درباره "تاریخ مطبوعات" ، "مطبوعات آذربایجان" و "تاریخ مطبوعات آذربایجان" است بازیابی کنیم. برای "تاریخ مطبوعات" باید شماره هایی را که هم عضو مجموعه "تاریخ" و هم عضو مجموعه "مطبوعات" است پیدا کرد و این همان ضرب منطقی این دو مجموعه است. بنابر آنچه گفتیم این مجموعه را می توان بسادگی پیدا کرد:

$$\{3, 23, 15, 456, 78, 89\} \cap \{11, 2, 12, 23, 94, 15, 456\} = \text{مطبوعات} \cap \text{تاریخ} \\ = \{23, 15, 456, 78\}$$

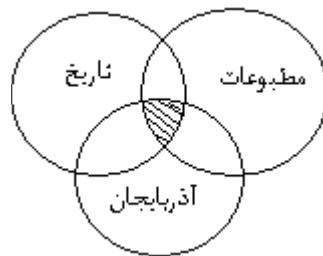
هم چنین مجموعه "مطبوعات آذربایجان" :

$$\{3, 23, 15, 456, 78, 89\} \cap \{11, 14, 456, 78, 89\} = \text{آذربایجان} \cap \text{مطبوعات} \\ = \{456, 78, 89\}$$

برای پیدا کردن شماره مدرکهای مربوط به "تاریخ مطبوعات آذربایجان" هم می توان حاصل ضرب منطقی سه مجموعه "تاریخ" و "مطبوعات" و "آذربایجان" را پیدا کرد و هم می توان حاصل ضرب مجموعه "تاریخ مطبوعات" را که پیدا کرده ایم، با مجموعه "آذربایجان" بدست آورد. ما روش دوم را بر می گزینیم و روش اول را به عهده خواننده می گذاریم:

$$\{11, 14, 456, 78, 89\} \cap \{23, 15, 456, 78, 89\} = \text{آذربایجان} \cap (\text{مطبوعات} \cap \text{تاریخ}) \\ = \{456, 78\}$$

نمودار حالت اخیر چنین است:

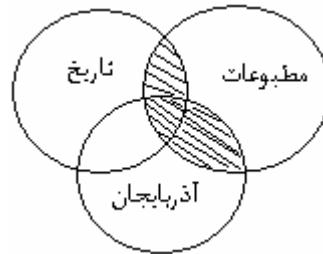


بخش هاشور خورده که مشترک میان هر سه دایره است مجموعه شماره مدرکهای مربوط به "تاریخ مطبوعات آذربایجان" را نشان می دهد

جمع منطقی

اگر پژوهنده ای بخواهد مدرکهای مربوط به تاریخ مطبوعات یا مطبوعات آذربایجان را بازیابی کند باید هم مدرکهای مربوط به تاریخ مطبوعات را پیدا کند و هم مدرکهای مربوط به مطبوعات آذربایجان را. البته این پژوهشنده به مدرکهای مربوط به تاریخ مطبوعات آذربایجان هم نیاز خواهد داشت. در اینجا بنابر آنچه گفتیم باید جمع منطقی دو مجموعه "تاریخ مطبوعات" و "مطبوعات آذربایجان" را پیدا کنیم. هر کدام از این دو مجموعه خود ضرب منطقی دو مجموعه هستند که آنها را در مثالهای بالا بدست آوردیم. روش محاسبه این جمع منطقی و نمودار آن چنین است:

$$\{456, 78, 89\} \cap \{23, 456, 78, 89\} = \{23, 456, 78\}$$



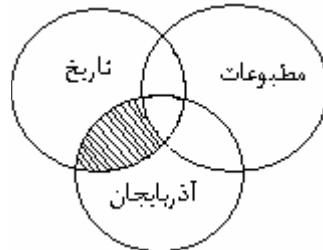
بخش هاشور خورده جمع منطقی خواسته شده است.

تفريق منطقی

اگر تنها به مدارکی درباره تاریخ آذربایجان و نه مطبوعات آن نیاز داشته باشیم نخست عضوهای مشترک دو مجموعه تاریخ و آذربایجان را در برگه های جدول می یابیم و از این مجموعه عضوهایی (شماره هایی) را که در برگه جدول مطبوعات هم آمده است حذف می کنیم. روش نوشتمن و نمودار این عمل چنین است:

$$-\{78, 456, 15, 23, 3, 89\} = \{78, 456, 11\} - \{78, 456, 15, 23, 3\} = \{11\}$$

$\{78, 456, 15, 94, 23, 12, 2, 11\} \cap \{89, 78, 456, 14, 11\}$ = مطبوعات - (آذربایجان و تاریخ)



البته در کار با برگه های جدول جمع و ضرب و تفریق شماره ها با چشم انجام میشود و هیچ نیازی به نوشتن مجموعه شماره ها باشکل مثالهای بالا نیست. اما کار عملی با برگه های جدول، مانند کارهای عملی دیگر، مبتنی بر اصلهایی نظری است که هم مبنای اعتبار آنند و هم دانستن آنها به رشد و تکامل آن کمک می کند. واقع امر این است که کاربرد اصلی منطق بول در نظامهای ذخیره و بازیابی خودکار (کامپیوتری) است. در این نظامها پرسشهایی که از کامپیوتر میشود همه در قالب عبارتهایی است که شامل علامتهای منطق بول هستند.

معرفی چند مأخذ

برای آشنایی با جبر بول خواندن کتاب زیر که ساده و در عین حال دقیق نوشته شده است توصیه می شود. نویسنده این کتاب از صاحب نظران منطق جدید است.

1- E. Mendelson, Boolean Algebra and Switching Circuits, Schaum's Outline Series, N.Y. 1970

برای مطالعه جدی تر و برای آشنایان با منطق و ریاضیات جدید فصل جبر بول در کتاب زیر از بهترین نوشته ها در این موضوع است.

2- J. Bell & M. Machover, A Course in Mathematical Logic, N.Y., Oxford, 1977

برای آنان که از نمایه سازی همراه هیچ گونه آگاهی ندارند مقاله زیر توصیه می شود:
نورالله مرادی. "نمایه سازی همارا". نامه انجمن کتابداران، دوره ۹ (بهار ۱۳۵۵) ص ۸۳-۱۰۷، (تابستان ۱۳۵۵) ص ۲۷۲-۲۹۶.