صفحه اصلی

اطلاع رسانی
نشریه فنی مرکز اطلاعات ومدارک علمی ایران
دوره دوازدهم؛ شماره ۳

قواعد استنتاج [۱] و فرمول بندی پرسش [۲] در کاوش رایانه ای

نوشته: محمدرضا داورپناه
دانشکده علوم تربیتی دانشگاه فردوسی مشهد

کلیدواژه ها
فرمول بندی پرسش کاوش رایانه ای عملگرهای بولی توصیفگر بانک اطلاعاتی

چکیده

برای بازیابی اطلاعات، باید پرسش ها فرمول بندی شوند. این کار باید با توجه به برنامه رایانه و با استفاده از عملگرهای منطقی بول صورت گیرد. کاوشگران همواره باید کار کاوش را با اخص ترین توصیفگرها فرمول بندی کنند. از آنجا که مرتبط کردن و ترکیب مفاهیم دارای اساس منطقی است، آشنایی با قواعد استنتاج منطقی ضروری است.

نقش فرمول بندی پرسش در فعالیت های بازیابی اطلاعات به اندازه نمایه سازی مهم است، و حتی از نظر برخی متخصصان اطلاع رسانی در بعضی موارد اهمیت آن از نمایه سازی هم بیشتر است. فرمول بندی پرسش به طور کلی شامل انتخاب و گزینش مفاهیم، نوع ارتباط بین مفاهیم و ترکیب آنها در قالب یک عبارت، تعیین حدود و ارزش مفاهیم به عنوان ملاک و میزان کاوش، و برقراری ارتباط بین زبان فرمولی پرسش و زبان سیستم می باشد. در نوع ساده تر و مثلاً ((در یک کاوش دستی، فرمول بندی پرسش شامل انتخاب توصیفگرهای موضوعی و در صورت امکان ترکیب آن با عملگرهای منطقی AND، OR و NOT است. کیفیت فرمول بندی پرسش به ساختار طبقه بندی واژگان و به ویژه ساختار زبان نمایه، توان، قابلیت و پیچیدگی زبان پرسش، اطلاعات به کار گرفته شده، مقدار زمان صرف شده، وجود قواعد و مقررات هدایتی به منظور فرمول بندی پرسش، زمینه و قابلیت کاوشگر، و روش کاوش از نظر امکان تقابل و بازخورد بستگی دارد.

فرمول بندی پرسش، تجزیه و تحلیلی ظریف را می طلبد، چرا که ساختار پایگاه اطلاعاتی می‌تواند به فرمول بندی پرسش بسیار کمک کند. پایگاه های اطلاعاتی معمولاً انواع مفاهیم را بر اساس ویژگی هایشان طبقه بندی می کنند تا از طریق این قالب بندی، فکر کاوشگر متناسب با نظام ذخیره داده ها در سیستم، به درستی جهت داده شود. ((مطالب مربوط به چیزهای تازه وقتی برای ما قابل درک می شود که

بتوانیم ارتباط آنها را با آنچه که در ذهن داریم بیابیم و به عبارت دیگر آنها را رده بندی کنیم. پس اگر، سیستمی که توسط آن به این دانش تازه دست می یابیم درباره رابطه ها اطلاع نادرستی بدهد نه تنها کمکی به ما نمی کند بلکه مانع پیشرفت تحقیقات ما نیز می شود [۳].)) در طبقه بندی های تحلیلی ـ ترکیبی، اصطلاحات غالباً بسیط هستند تا در هنگام نیاز، بتوان آزادانه آنها را ترکیب کرد. از این رو جز در موارد خاص، نیازی به ارائه توصیفگرهای مرکب در قالب های از پیش ساخته نیست. همین ویژگی امکان می دهد که حتی بسیاری از واژه های مرکب را هم بتوان به عوامل معنایی یا نحوی آنها تجزیه کرد و با ارائه مفاهیم یا اصطلاحات بسیط، نه تنها جنبه های مهم یک مفهوم را به روشنی بیان کرد، بلکه موجبات بررسی و تعیین روابط مفاهیم و کاوش با هر نوع ترکیبی از مفاهیم را نیز بخوبی فراهم آورد.

به عکس، در رده بندی های سنتی دیویی و کنگره هر توصیفگر علی الاصول نماینده یک موضوع کاملاً خاص و منحصر به فرد است؛ هر مفهوم فقط تابع یک مفهوم عام تر می باشد و در واقع نظام، تک سلسله ای است. ولی در نظام های تحلیلی ـ ترکیبی، هر یک از مفاهیم با تک تک مفاهیم دیگر سنجیده می شود تا مشخص شود که آیا با هم رابطه مرتبه ای دارند یا خیر، و اگر دارند مرتبه وابستگی آنها چیست. ممکن است برای یک مفهوم بیش از یک مفهوم عام تر وجود داشته باشد که بدین ترتیب، طرح رده بندی چند سلسله ای فراهم می آید. [۴] بنابراین، از آنجا که در طبقه‌بندی‌های تحلیلی ـ ترکیبی موضوع ها تحلیل می شوند و توالی خطی بر این طبقه بندی ها حکم فرما نیست و رابطه ای چند سلسله ای فراهم می آورند، به نظر می رسد در راهنمایی جستجوگر در تجزیه و تحلیل موضوع پرسش و گزینش توصیفگرهای موضوعی بسیار مفید و موثر باشند. همچنین، اگر طبقه ها به شیوه های مختلف از جمله رده ای، الفبایی یا درختی نمایش داده شوند کاوشگر یا استفاده کننده می‌تواند وابستگی و ارتباط بین مفاهیم را به روشنی تشخیص دهد و مفاهیم و توصیفگرهای مورد نیاز خود را با توجه به سطح اعم و اخص بودن و نیز ارتباط سلسله مراتبی بین آنها به درستی و دقت انتخاب کند. البته ساختار زبان نمایه باید یک سیستم حمایتی باشد و نباید جانشین تفکر خلاق و تجزیه و تحلیل منطقی کاوشگر شود.

منطق خاص فرمول بندی پرسش باید بر اساس قضاوتی درست، آگاهی از برنامه رایانه، و با استفاده از عملگرهای منطق بولی ترسیم بشود. کاوشگر خوب کسی است که همه امکاناتی را که در مکانیزم بازیابی فراهم آمده برای بسط و گسترش منطق پرسش به کار گیرد. به منظور بازیابی مطلوب، کاوشگر باید کاوش را در ابتدا با اخص ترین فرمول بندی ممکن شروع کند و در صورت عدم موفقیت، پرسش را مرحله به مرحله عام تر نماید. به کارگیری این شیوه خود به فهم و درک دقیق پرسش و تجزیه و تحلیل درست نیازمند است.

منطق رایانه

درباره منطق رایانه در متون تخصصی مطالب مبسوطی بیان شده، ولی از آنجا که در متون مربوط به کتابداری و اطلاع رسانی در خصوص اساس و چگونگی عملکرد آن کمتر سخن به میان آمده و ادامه این بحث نیز موکول به شناخت خواننده از این مبانی است، از این رو مقدمتاً به توضیح آن می‌پردازیم.

اساس منطق رایانه ها را جبر بول که توسط جرج بول ریاضی دان انگلیسی قرن نوزدهم ابداع شد، تشکیل می دهد. منطق علامتی بول سال های متمادی بعد از اکتشاف مورد توجه قرار نگرفت، ولی امروزه منطق ریاضی یکی از شاخه های مهم ریاضی است و زبان منطق، زبان محاسبه در رایانه است.

عملیات منطقی و مرتبط ساختن و ترکیب مفاهیم جداگانه با یکدیگر از طریق منطق بول با عملگرهای AND (و)، OR (یا) NOT (نه) صورت می گیرد.

متغیرهای منطق بول نیز فقط دو حالت دارند: یا درست ند یا نادرست، که در متون مربوطه برای سهولت کار، آنها را با "F"(FALSE) و "T"(TRUE) یا با ((د)) (درست) و ((ن)) (نادرست) نشان می دهند. اما از آن جا که کوچک ترین جزء اطلاعاتی در رایانه که بیت (BIT) نام دارد نشان دهنده دو حالت است و یکی از این حالات به ((۰)) و دیگری به ((۱)) تعبیر می شود، بنابراین به منظور نمایش جدول ارزشی عملگرها که در واقع بیانگر چگونگی کار عملگر در رابطه هاست، به جای (T) و (F) از صفر و یک استفاده می شود.

۱- ترکیب عطفی

رابط این نوع ترکیب همان ((و)) (یا "AND") در زبان محاوره ای است که عاطف نیز می‌نامند. استفاده از AND در ترکیب مفاهیم، کاوش را محدود و اخص می سازد و در واقع یک ضرب منطقی است که در منطق با علامت ((.)) نمایش داده می شود. در این نوع ترکیب، نتیجه یا خروجی در صورتی صادق است که هر دو مولفه صادق، ((۱)) باشند و جدول ارزش آن بر اساس کلیه حالات ممکن با استفاده از فرمول ۲^n تعیین می شود. چنانچه فقط یک مولفه داشته باشیم، بر اساس این فرمول $۲^۱ = ۲$. بنابراین، یک مولفه دو حالت دارد: یا درست است یا نادرست، یا ((۰)) است یا ((۱)). ولی، دو مولفه با هم چهار حالت دارند (زیرا $۲^۲ = ۴$): امکان دارد هر دو ((۰))، هر دو ((۱)). اولی ((۱)) دومی ((۰))، اولی ((۱)) و دومی ((۱)) باشد.

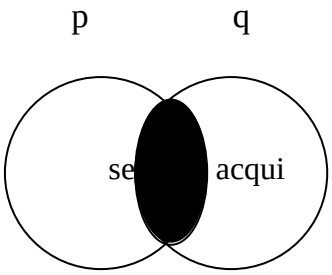
چنانچه در یک کاوش رایانه ای بخواهیم تمام رکوردهای محتوی دو واژه "Serials" و "acquisition" بازیابی گردند و بدین منظور این واژه ها را به صورت "Serials AND acquisition" فرموله کنیم، با توجه به جدول ارزشی AND فقط رکوردهایی بازیابی و نمایش داده می شوند که هم شامل واژه "serials" و هم شامل واژه "acquisition" باشند؛ یعنی هر دو ((۱)) باشند. رکوردهایی که فقط محتوی یکی از واژه ها باشد با این فرمول بندی بازیابی نخواهد شد.

جدول ۱ چگونگی ترکیب مفاهیم را نمایش می دهد. به جهت رعایت اختصار به جای مفهوم واژه ها از نمادهای((p)) و ((q)) در جدول استفاده شده است.

جدول ۱: جدول عملگر "AND"

ورودی		خروجی
p	q	p.q
0	0	0
0	1	0
1	0	0
1	1	1

چون ترکیب مفاهیم با استفاده از عملگرهای بول معمولاً به وسیله نمودارهای ون نیز نشان داده می‌شود، نمودار ۱ فرمول مورد کاوش را در قالب نمودار ون نمایش می دهد. قسمت هاشور خورده، جواب کاوش است.



نمودار ۱: نمودار ″ون ″برای عملگر″AND″

۲- ترکیب فصلی

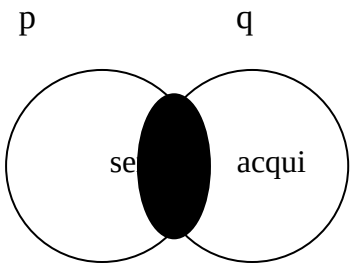
یکی از حروفی که در زبان محاوره ای به عنوان رابط بین جملات استعمال می شود ((یا)) (OR) است که در ترکیب منطقی باید آن را با دقت به کار برد. در زبان محاوره از جمله ((هوا برفی است یا خورشید در آسمان دیده می شود))، بیشتر منظورمان این است که یکی از دو وضع واقع شده و هر دو نیست. اما در منطق، ترکیب منطقی یا ″OR″ در واقع جمع منطقی است که شامل هر دو وضعیت است و در منطق با علامت ″v″ نشان داده می شود.

در این نوع ترکیب، خروجی در صورتی صادق یا ((۱)) است که هیچ کدام از دو مولفه ((q)) و ((p)) کاذب با ″۰″ نباشند. به عبارت دیگر، چنانچه هر دو مولفه صادق یا یکی از مولفه ها صادق باشد، نتیجه صادق است.

برای روشن تر شدن مطلب، چنانچه عبارت مثال قبلی به صورت ″Serials or Acquisition″ فرموله و به رایانه داده شود، با عنایت به نتیجه جدول ارزشی، ارجاعاتی بازیابی می‌شود که محتوی هر دو واژه یا هر یک از این دو واژه، اعم از ″Serials″ یا ″Acquisition″ باشند و فقط ارجاعاتی بازیابی نمی شود که شامل هیچیک از این دو واژه نباشد. (جدول و نمودار ۲)

جدول شماره ۲: جدول عملگر ″OR″

خروجی	ورودی	
pvq	q	p
0	0	0
1	1	0
1	0	1
1	1	1



نمودار ۲: نمودار″ ون ″ برای عملگر″OR″

۳- عملگر منطقی NOT

عملگر منطقی NOT معکوس ساز است؛ یعنی ((۰)) را به ((۱)) یا ((۱)) را به ((۰)) تبدیل می کند و در منطق با علامت ((~)) نمایش داده می شود. این عملگر برای مستثنی کردن یک مفهوم بخصوص، یا یک مفهوم خاص از مفهوم عام تر مورد استفاده قرار می گیرد و در واقع تفاضل منطقی است. جدول عملگر NOT به صورت ذیر است:

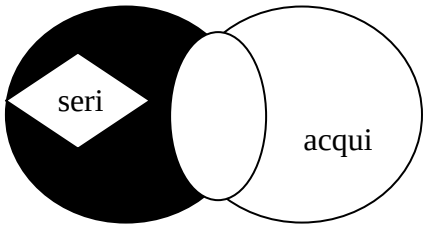
جدول ۲: جدول عملگر ((NOT))

ورودی	خروجی
q	p
0	1
1	0

البته جدول ارزشی "NOT" در ترکیب با عملگر دیگر مثل "AND" شکل دیگری پیدا می کند؛ مثلاً جدول ارزشی (p.q) ~ چنین می شود:

~	(p	0	q)
0	1	1	1
1	1	0	0
1	0	0	1
1	0	0	0

اگر مثال قبلی به صورت Serials Not Acquisition به رایانه داده شود، نتیجه کاوش برابر نمودار شماره ۳ یک مجموعه کوچک تر خواهد بود.



نمودار ۳: نمودار "ون" برای عملگر "NOT"

((روش تفاضل منطقی بیشتر مناسب گروه هایی از توصیفگرهاست که مانعه الجمع اند؛ یعنی مواردی که فقط یک توصیفگر از آن گروه برای نمایه سازی هر سند برگزیده شده است. ((تاریخ)) سند مثالی از این گونه توصیفگرهای مانعه الجمع می باشد. اگر توصیفگر ((کتابخانه های مدارس)) برای تعدادی از اسناد برگزیده شده باشد و بخواهیم این مدارک را بر مبنای تاریخ انتشار آنها گزینش کنیم، کاوش کتابخانه های مدارس در غیر از مواردی که پیش از ۱۳۶۵ منتشر شده موجب می شود که اسناد منتشر شده مربوط به کتابخانه های مدارس در پیش از ۱۳۶۵ بازیابی نشود [۵].)) (نمودار ۴)

نمودار ۴

عملگر منطقی "NOT" کمتر مورد استفاده قرار می گیرد، زیرا در مواردی که دو توصیفگر را بر این اساس فرموله می کنیم، مستثنی کردن و تفاضل منطقی باعث می شود مدارکی که اطلاعات مفیدی دارند، یعنی هر دو توصیفگر را با هم مورد توجه قرار داده اند، بازیابی نشوند.

به کارگیری قواعد استنتاج در فرمول بندی پرسش

در زبان محاوره ای ممکن است یک کلمه یا یک جمله در متن های مختلف یا در شرایط مختلف، حتی در بیان های متفاوت، معانی متفاوتی داشته باشد. مثلا جمله ((حالا ساعت ۹ صبح است)) در بیان عادی، یک جمله ساده است که ساعت ۹ صبح را خبر می دهد. ولی فرض کنید که شما به دوستتان قول داده اید ساعت ۹ صبح در کتابخانه خواهید بود، و اتفاقاً ساعت ۱۰ صبح به کتابخانه می رسید. دوستتان که در کتابخانه منتظر شما بوده با دیدن شما با ناراحتی می پرسد ((حالا ساعت ۹ صبح است؟)) معنی این جمله در اینجا مسلماً با معنی قبلی آن فرق دارد. اما زبان فرمولی، باید دقیق باشد و ما از آزادی هایی که در زبان محاوره ای داریم برخوردار نیستیم. ((استدلال منطقی عبارت است از حرکت از معلوم به مجهول، و در منطق جدید بر طبق قواعد و اصولی انجام می پذیرد که این قواعد و اصول به ((قواعد استنتاج)) معروف ند. به کارگیری این قواعد باعث سهولت در کار استدلال می شود [۶].))

اما آنچه در اینجا در نظر است به کارگیری این قواعد برای فرمول بندی پرسش در کاوش رایانه ای است. مرتبط ساختن و ترکیب مفاهیم اساس منطقی دارد و پرسشها بنا بر اصول منطقی باید فرمول بندی شوند. شناخت قواعد استنتاج منطقی کاوشگر را در فرمول بندی کردن پرسش یاری می‌کند و او را با شیوه های مختلف فرمول بندی آشنا می سازد. بر اساس این اصول روشن می شود که دستیابی به نتیجه واحد، با فرمول بندیهای معادل امکان پذیر است. یعنی دو نوع فرمول بندی منطبق با اصول استنتاجی، کاوشگر را به یک نتیجه واحد رهنمون می سازد.

برخی از این قواعد که به نظر می رسد با توجه به عملگرهای منطق بول (NOT، OR، AND) در کار کاوش رایانه ای اطلاعات بیشتر مورد استفاده قرار گیرند عبارت ند از:

جابه جایی (Commutation)	Comm	pvq=qvp p.q=q.p
شرکت پذیری (Association)	Assoc	pv(qvr)=(pvq)vr p. (q.r)=(p.q).r
بخش پذیری (Distribution)	Dist	pv(q.r)=(pvq).(pvr) p.(qvr)=(p.q)v(p.r)
قوانین دمرگان (De Morgan)	Dem	~(pvq)= ~p. ~q ~(p.q)= ~pv~q

مبنای تنظیم قواعد استنتاجی بالا که به ((قواعد جایگذاری)) [۷] معروف اند هم ارزهای منطقی است. برای نمایش هم ارزی از علامت ((≡)) استفاده می شود. در منطق در صورتی که دو قضیه دارای جدول ارزشی یکسانی باشند آن دو را نسبت به هم هم ارز می خوانند؛ مثل قواعد بالا که قضایای دو طرف علامت ((≡)) نسبت به هم هم ارز هستند [۸].

مسلم است که به منظور کاوش اطلاعات در رایانه، باید پیش از برقراری تماس با صفحه کلید مجموعه ای از عملیات انجام پذیرد. نهایی ترین این مراحل فرمول بندی است؛ یعنی پرسش در قالب فرمولی که برای رایانه قابل درک و در راستای نیل به نتیجه دلخواه پرسشگر باشد تنظیم و پس از طی مراحل بعدی به رایانه داده می شود.

تفاوت فرمول های موازی و هم ارز مدت زمان کاوش آنها توسط رایانه است. بویژه در فرمول‌هایی که از پرانتز برای جلوگیری از ابهام استفاده می شود، عملیات کاوش رایانه ای بسته به تعداد پرانتزها کندتر و وقت بیشتری صرف کاوش می گردد. فرضاً محققی در پی مقالاتی درباره ((علل افزایش فشارخون بعد از جراحی کلیه)) است. همانطور که ملاحظه می شود موضوع مورد نظر سه کلیدواژه: ((کلیه)) (Kidnay)، ((فشارخون)) (hypertension) و جراحی (surgery) را شامل می‌شود. به منظور ملموس تر بودن مطلب، این سه کلیدواژه را در فرمول بندی های کاوش با هم ترکیب می کنیم و با توجه به نمادهای اصول استنتاج، کلیدواژه ها را برای تسهیل درک ارتباط بین اصول استنتاجی و فرمول بندی کاوش به صورت زیر برمی گزینیم.

P=kidney
q=hypertension
r=surgery

برای به دست آوردن نتایج عینی و مقایسه عملی نتایج حاصله، هر یک از فرمول بندی ها در بانک اطلاعاتی ((مدلاین)) (در فاصله زمانی ژانویه تا سپتامبر سال ۱۹۹۲) کاوش شدند که چگونگی فرمول‌بندی ها و نتایج به دست آمده از هر یک در زیر از نظرتان می گذرد.

الف. اصل جابه جایی

این اصل بدیهی ترین اصل است و همان گونه که از نامگذاری آن پیداست در حقیقت کلیه واژه های دو طرف رابطه با یکدیگر جابه جا می شوند:

1) pvq ≡ qvp

↓ تعداد رکورد بازیابی شده

(۱۸۹۴۲) کلیه OR فشار خون ≡ (۱۸۹۴۲) فشار خون OR کلیه

2) p.q ≡ q.p

(۹۵۲) کلیه AND فشارخون ≡ (۹۵۲) فشارخون AND کلیه

همانطور که ملاحظه می شود تعداد رکوردهای بازیابی شده در هر یک از فرمولهای معادل یکسان است و با یکدیگر تفاوتی ندارند. به عبارت دیگر، هر فرمول نتیجه واحدی را به دست داده است.

ب. اصل شرکت پذیری

1) pv(qvr) ≡ (pvq)vr

≡ (۵۹۱۴۵) (جراحی OR فشارخون) OR کلیه

(۵۹۱۴۵) جراحی OR (فشارخون OR کلیه)

2) p.(q.r) ≡ (p.q).r

≡ (۸۳) (جراحی AND فشارخون) AND کلیه

(۸۳) جراحی AND (فشارخون AND کلیه)

ج. اصل بخش پذیری

بنابراین اصل هر یک از فرمول ها به فرمول های معادل دیگری قابل تقسیم خواهند بود.

1) pv(q.r) ≡ (pvq).(pvr)

≡ (۱۳۲۲۵) (جراحی AND فشارخون) OR کلیه

(۱۳۲۲۵) (جراحی OR کلیه) AND (فشارخون AND کلیه)

2) p.(qvr) ≡ (p.q)v(p.r)

≡ (۲۳۸۰) (جراحی OR فشارخون) AND کلیه

(۲۳۸۰) (جراحی AND کلیه) OR (فشارخون AND کلیه)

در فرمول بندی های معادل، تعداد پرانتزها بیشتر شده که در این حالت مدت زمان کاوش دو فرمول با یکدیگر متفاوت خواهد بود، اما نتایج به دست آمده از هر دو فرمول موازی یکسان است.

د. قوانین دمرگان

از آنجا که نتایج به دست آمده از این نوع فرمول بندی صفر است، معمولاً این نوع فرمول‌بندی به تنهایی برای کاوش به کار گرفته نمی شود، ولی در ترکیب با سایر فرمول بندی ها قطعاً مورد استفاده دارد و شناخت آن برای کاوشگر بی فایده نخواهد بود:

1) ~ (p.q) ≡ ~p. ~q

NOT (کلیه OR (فشارخون) (۰) ≡ NOT کلیه AND NOT فشارخون (۰)

2) ~ (p.q) ≡ ~pv ~q

NOT (کلیه AND (فشارخون) (۰) ≡ NOT کلیه OR NOT فشارخون (۰)

با نظری اجمالی به فرمول ها و تعداد رکوردهای به دست آمده در هر یک، در می یابید که فرمول های معادل عملاً نتایج یکسانی به دست داده اند. همچنین با مقایسه نتایج هر یک از فرمول‌های اصل جابجایی و اصل شرکت پذیری با هم، مفهوم اخص سازی عملگر AND و گسترش‌پذیری عملگر OR به روشنی مشخص می شود.

ارجاعات و یادداشت ها:

1) Rules of inference

2) Query formulation

۳- فوسکت، دی. جی. ((مسائل نمایه سازی نشریات علوم اجتماعی)). ترجمه مهین دستمالچی، نشریه فنی مرکز اسناد و مدارک علمی، دوره ۳ (شماره ۳ و ۴)، ۱۳۵۴.

۴- حریری، مهرانگیز. ((اصطلاحنامه در نظام ذخیره و بازیابی اطلاعات)). اطلاع رسانی، نشریه فنی مرکز اسناد و مدارک علمی، دوره ۶ (شماره ۱ و ۲) ۱۳۶۱، دوره ۷ (شماره ۱) ۱۳۶۱.
۵- جهودا، جرالد، براوناگل، جودیث شیک. مبانی مرجع: فنون پرسش کاوی و پاسخ یابی از منابع کتابخانه. ترجمه و تالیف محمد حسین دیانی، عبدالحمید معرف زاده، با مقدمه و ویرایش اسدالله آزاد، مشهد، آستان قدس رضوی، ۱۳۷۰.
۶- مصاحب، غلامحسین، مدخل منطق صورت. انتشارات حکمت، چاپ دوم، ۱۳۶۶.
7- Rules of replacement
۸..مصاحب، غلامحسین، همان.