

IS specifikavimo kalbų vidinės kokybės atributų agregavimo ypatumai

Jelena GASPEROVIČ, Albertas ČAPLINSKAS (MII)

el. paštas: j.gasperovic@algoritmusistemas.lt, alcapl@ktl.mii.lt

Reziumė. Straipsnyje nagrinėjami IS specifikavimo kalbų vidinės kokybės atributų agregavimo ypatumai. Aptariami agregavimo operatoriai, išskiriamos agreguojamų kalbos vidinės kokybės atributų grupės. Pagrindinis nagrinėjimų tikslas – kiekvieno tipo atributų grupei pasiūlyti tokių kokybės atributų reikšmių agregavimo metodą, kuri panaudojus nebūtų iškreipta pirminių matavimų prasmė.

Raktiniai žodžiai: IS specifikavimo kalba, vidinė kokybė, taksonominė hierarchija, agregavimo metodas.

Įvadas

Kuriamų produktų ar teikiamų paslaugų kokybė vertinama praktiškai visose veiklos srityse. Dažnai yra vertinama taip pat ir naudojamų įrankių ar priemonių kokybė, jų tinkamumas konkrečiam darbui atlikti. Tai daroma ir informacinių sistemų inžinerijoje, kur yra vertinamas naudojamų priemonių, pavyzdžiui, IS specifikavimo kalbų tinkamumas konkrečioms informacinėms sistemoms kurti. Tačiau, nepaisant tokios didelės kokybės svarbos, kol kas ne tik nėra sukurta bendra kokybės vertinimo teorija, bet net ir pati kokybės sąvoka nėra griežtai apibrėžta. Tarptautiniai kokybės standartai, pavyzdžiui, ISO 9000 standartų šeima, bei jų pagrindu sukurti specializuoti kokybės standartai, pavyzdžiui, programinės įrangos kokybės standartas ISO 9126 [7] taip pat nepateikia nei išsamios kokybės vertinimo sampratos, nei konkrečių kokybės vertinimo metodikų. Taip yra todėl, kad skirtingose veiklos srityse kokybė yra suprantama labai skirtingai ir todėl kiekvienai veiklos sričiai neišvengiamai tenka kurti savus kokybės vertinimo metodus.

Savo ankstesniuose darbuose [3], [4] pasiūlėme IS specifikavimo kalbos kokybės sampratą ir bendrą tos kokybės vertinimo metodiką. Pasiūlymų esmę trumpai galima apibūdinti šitaip. Visų pirma yra atskiriama kalbos panaudojimo kokybė ir jos vidinė kokybė. Panaudojimo kokybė nusako kalbos tinkamumą konkrečiam projektui. Ji vertinama atsižvelgiant į tam projektui svarbias kalbos savybes (vertinimo kriterijus) ir tų kriterijų prioritetus. Iš esmės, tenka spręsti daugiakriterinio sprendimų priėmimo uždavinį. Šis uždavinys yra sprendžiamas panaudojant kalbos kokybės atributų medį, aprašantį vidinę vertinamos kalbos kokybę. Vidinė kokybė nusako santykinę kalbos „gerumą“ lyginant ją su kitomis kalbomis arba, kitaip tariant, tos kalbos savybių išsamumo laipsnį. Šį laipsnį pasiūlyta [3] matuoti tikimybe, kad atitinkama kalbos savybė bus pakankama bet kuriai teoriškai įmanomai informacinei sistemai specifikuoti, ir matavimui panaudoti reprezentatyviųjų pavyzdžių bibliotekas. Kiekvienai IS speci-

fikavimo kalbai jos vidinės kokybės vertinimas turi būti atliktas vieną kartą. Šio vertinimo rezultatai panaudojami daug kartų, vertinant kalbos tinkamumą konkrečioms informacinėms sistemoms specifikuoti. Pažymėtina, kad aptartasis IS specifikuojimo kalbų kokybės vertinimo būdas yra originalus ir iš esmės skiriasi nuo kitų autorių siūlytų būdų, apžvelgtų darbuose [1], [2].

Atlikus išsamią svarbiausių IS specifikuojimo kalbų ir jų panaudojimo ypatumų analizę [1], [2] bei atsižvelgiant į kokybės vertinimo patirtį gretimose srityse (programavimo kalbų, koncepcinių modelių kokybė ir kt.), buvo nustatyta [3] kad IS specifikuojimo kalbų kokybė aprašoma taksonomine hierarchija $x = (x_1(x_{11}(x_{111}, x_{112}, x_{113}), x_{12}(x_{121}(x_{1211}, x_{1212}, x_{1213}, x_{1214}), x_{122}(x_{1221}, x_{1222}, x_{1223}, x_{1224}(x_{12241}, x_{12242}))), x_2(x_{21}(x_{211}, x_{212}, x_{213}, x_{214}(x_{2141}, x_{2142}, x_{2143}), x_{215}, x_{216}, x_{217}), x_{22}), x_3(x_{31}(x_{311}, x_{312}, x_{313}, x_{314}, x_{315}, x_{316}, x_{317}), x_{32}(x_{321}, x_{322}, x_{323}, x_{324}, x_{325})), x_4(x_{41}, x_{42}, x_{43}, x_{44}, x_{45}, x_{46}, x_{47}(x_{471}(x_{4711}, x_{4712}), x_{472})))$. Apatinių hierarchijos lygmenį sudaro 43 atributai, tačiau kai kurie iš jų kartojasi skirtingose šakose. Skirtingų (t.y. savarankiškai matuojamų) atributų yra 34. Kadangi kalbos tinkamumui konkrečiam projektui vertinti gali prireikti bet kurio lygmens vidinės kokybės atributų įverčių, tai turi būti suskaičiuotos visų vidinės kokybės atributų reikšmės. Kitaip tariant, žemiausiojo lygmens atributai turi būti agreguojami pagal atitinkamas grupes ir taip kylama taksonominėje hierarchijoje iki pat jos viršutinio taško. Šis uždavinys nėra trivialus, nes tas pats agregavimo metodas visiems atributams agreguoti negali būti panaudotas. Agregavimo metodas priklauso nuo agreguojamų atributų pobūdžio ir jų tarpusavio sąryšių. Netinkamai parinkus agregavimo metodą, yra iškreipiama pirminių matavimų prasmė. Pavyzdžiui, agreguojant funkcionalumo ir patikimumo savybes galima imti aritmetinį vidurkį, nes šios savybės yra nepriklausomos, ortogonalios viena kitai. Tačiau toks agregavimo būdas visiškai netiktų agreguojant kalbos universalumo ir jos adaptyvumo savybes, nes funkcionalumo požįūriu jos gali tam tikru mastu dubliuoti viena kitą. Be to, mūsų atveju, dar reikia atsižvelgti į tikimybinę vidinės kokybės charakteristikų vertinimo pobūdį.

Šio straipsnio tikslas aptarti, kokios skirtingos atributų priklausomybės susidaro nagrinėjamoje taksonominėje atributų hierarchijoje ir kokie agregavimo metodai taikantini kokybės atributų reikšmėms agreguoti kiekvienu iš tų atvejų.

Agregavimo operatoriai

Visų pirma trumpai priminsime kai kuriuos agregavimo teorijos teiginius. Teoriniuose nagrinėjimuose paprastai yra daroma prielaida [6], kad tiek agreguojamos reikšmės, tiek ir agregavimo rezultatas priklauso kokiam nors baigtiniam intervalui, dažniausiai, intervalui $[0,1]$, nes, panaudojus atitinkamą tiesinę transformaciją, bet kurią baigtinį intervalą galima transformuoti į intervalą $[0,1]$. Tuomet agregavimo operatorius apibrėžiamas kaip funkcija, atvaizduojanti duotą reikšmių rinkinį $(x_1, x_2, \dots, x_n \mid x_i \in [0, 1])$ į reikšmę $y \in [0, 1]$. Agregavimo operatorius [5] turi būti monotoniškas, agreguojant vieną reikšmę kaip rezultatą gauti ją pačią (tapatumo savybė) ir tenkinti vadinamąsias kraštines sąlygas $\text{Aggreg}(0, \dots, 0) = 0$, $\text{Aggreg}(1, \dots, 1) = 1$. Iš šių savybių seka operatoriaus tolydumas bet kurio kintamojo atžvilgiu, asociatyvumas, komutatyvumas, bisimetriškumas ir idempotentiškumas. Mūsų nagrinėjamo už-

davinio požiūriu svarbios yra trys agregavimo operatorių kategorijos: konjunktyvieji operatoriai, dizjunktyvieji operatoriai ir operatoriai su svoriniais koeficientais.

Konjunktyvieji operatoriai – tai operatoriai, duodantys didelę reikšmę (t.y. artimą 1 reikšmę) tik tai tuomet, kuomet visos agreguojamos reikšmės taip pat yra didelės. Tai tam tikras loginio operatoriaus „IR“ analogas. Konjunktyvieji operatoriai tinka nepriklausomoms (ortogonalioms) savybėms agreguoti. Svarbi konjunktyviųjų operatorių klasė yra vadinamosios t -normos [6] (t -normas žymėsime raide T). Jos tenkina reikalavimą $T(x_1, x_2, \dots, x_n) \leq \min(x_1, x_2, \dots, x_n)$. Paprasčiausi intervale $[0,1]$ apibrėžtos dvivietės t -normos pavyzdžiai yra minimumas $\min(x_1, x_2)$ ir sandauga $x_1 x_2$.

Dizjunktyvieji operatoriai – tai konjunktyviesiems operatoriams dualūs operatoriai. Jie duoda didelę reikšmę tuomet, kuomet bent viena iš agreguojamų reikšmių yra didelė. Kitaip tariant, tai tam tikras loginio operatoriaus „ARBA“ analogas. Dizjunktyvieji operatoriai tinka viena kitą kompensuojančioms arba papildančioms savybėms agreguoti. Svarbi dizjunktyviųjų operatorių klasė yra vadinamosios t -konormos [6] (t -konormas žymėsime ženklu \perp). Jos tenkina reikalavimą $\perp(x_1, x_2, \dots, x_n) \geq \max(x_1, x_2, \dots, x_n)$. Paprasčiausi intervale $[0,1]$ apibrėžtos dvivietės t -konormos pavyzdžiai yra maksimumas $\max(x_1, x_2)$ ir tikimybių sudėtis $x_1 + x_2 - x_1 x_2$.

Operatoriai su svoriniais koeficientais naudojami tuomet, kuomet agreguojamosios reikšmės yra nevienodos svarbos. Paprasčiausias tokio operatoriaus pavyzdys yra vidurkis su svoriniais koeficientais. Svarbios tokių operatorių klasės yra svorinės t -normos T_w ir svorinės t -konormos \perp_w [8]. Tai atitinkami t -normų ir t -konormų apibendrinimai.

Agreguojamų atributų grupės

Panagrinėkime, kokie atributų tarpusavio sąryšiai galimi aptariamojoje taksonominėje hierarchijoje ir kokie agregavimo metodai taikytini kiekvienam iš galimų atvejų. Tegul x_i žymi kokybės atributą, $L(x_i)$ – tuo atributu aprašomą kalbos savybę, $\text{Orth}(x, y)$ – predikatas, igyjantis reikšmę „teisingas“ tuomet, kuomet atributai x ir y yra ortogonalūs (t.y. $L(x) \cap L(y) = 0$ ir savybės $L(x)$ ir $L(y)$ yra naudojamos skirtingoms informacinės sistemos savybėms specifikuoti), $\text{Supl}(x, y)$ – predikatas, igyjantis reikšmę „teisingas“ tuomet, kuomet atributas y aprašo savybę papildančią (išplečiančią) atributu x aprašomą savybę (t.y. $L(x) \cap L(y) = 0$, bet savybės $L(x)$ ir $L(y)$ yra naudojamos toms pačioms informacinės sistemos savybėms specifikuoti), $\text{Alt}(x, y)$ – predikatas, igyjantis reikšmę „teisingas“ tuomet, kuomet atributais x ir y aprašomos savybės yra alternatyvios (t.y. $L(x) \cap L(y) \neq 0$ ir savybės $L(x)$ ir $L(y)$ gali būti naudojamos toms pačioms informacinės sistemos savybėms specifikuoti). Naudojant šiuos žymėjimus, bet kurioje kalbos kokybę aprašančioje taksonominėje hierarchijoje teoriškai galimi tokie atributų tarpusavio sąryšiai:

- $\text{Ort}(x_1, x_2, \dots, x_n)$, t.y. visi agreguojamos grupės atributai yra ortogonalūs;
- $\text{Ort}(x_1, x_2, \dots, x_k)$, bet kuriam $i \in [k+1, n]$ egzistuoja bent vienas toks $j \in [1, k]$ arba $j \in [k+1, n]$, kad $\text{Supl}(x_j, x_i)$, ir bent vienam $i_1 \in [k+1, n]$ atitinkamas j_1 priklauso intervalui $[k+1, n]$, t.y. dalis agreguojamos grupės atributų yra ortogonalūs, kiti yra papildomi, jais aprašomos savybės tiesiogiai arba netiesiogiai

išplečia kuriuo nors baziniu atributu aprašomą savybę arba galbūt netgi kelias baziniais atributais aprašomas savybes;

- $\text{Alt}(x_1, x_2, \dots, x_n)$, t.y. visi agreguojamos grupės atributai yra alternatyvūs;
- $\text{Ort}(x_1, x_2, \dots, x_k)$, bet kuriam $i \in [k + 1, n]$ egzistuoja bent vienas toks $j \in [1, k]$, kad $\text{Supl}(x_j, x_i)$, ir egzistuoja bent viena tokia indeksų pora $l \in [k + 1, n]$ ir $r \in [k + 1, n]$, kad $\text{Alt}(x_l, x_r)$.

Vienok, nepažeidžiant bendrumo, taksonominę hierarchiją galima taip sukonstruoti, kad atributų sąryšiai supaprastėtų:

- atributai gali būti grupuojami taip, kad visi papildomi atributai tiesiogiai arba netiesiogiai papildytų tik vieną bazinį atributą;
- alternatyvūs atributai gali būti visuomet traktuojami kaip savarankiškos atributų grupės.

Toks taksonominės hierarchijos konstravimo būdas gerokai supaprastina antrąjį atvejį ir apskritai eliminuoja ketvirtąjį atvejį. Aptariamoji taksonominė hierarchija yra sukonstruota būtent šitaip. Tačiau, kita vertus, agreguojant atributus dar reikia atsižvelgti į atributais aprašomų savybių pobūdį. Savybė $L(x_i)$ gali būti tokia, kad be jos negalima apseiti specifikuojant jokią informacinę sistemą (t.y. tikimybė $p(x_i) = 1$) arba tokia, kad jos prireiks specifikuojant tik kai kurias sistemas (t.y. $p(x_i) = \alpha_i$, kur $\alpha_i < 1$). Be to, savybė $L(x_i)$ kalboje gali būti realizuota maksimaliai (t.y. tikimybė $q(x_i)$), kad jos pakaks bet kuriai sistemai, kurią specifikuojant jos prireiks, lygi 1) arba tik iš dalies (t.y. $q(x_i) = \beta_i$, kur $\beta_i < 1$). Įvertinus visus šiuos ypatumus, aptariamojoje taksonominėje hierarchijoje išskiriamos keturių tipų agreguojamų atributų grupės:

- $\text{Ort}(x_1, x_2, \dots, x_n)$ ir $p(x_i) = 1$) bet kuriam $i \in [1, n]$;
- $\text{Ort}(x_1, x_2, \dots, x_n)$ ir egzistuoja bent vienas toks $i \in [1, n]$, kad $q(x_i) = \beta_i$, kur $\beta_i < 1$;
- $\text{Supl}(x_1, x_2), \text{Supl}(x_2, x_3), \dots, \text{Supl}(x_{n-1}, x_n)$, $p(x_1) = 1$ ir $p(x_i) = \alpha_i$, kur $\alpha_i < 1$, bet kuriam $i > 1$;
- $\text{Alt}(x_1, x_2, \dots, x_n)$, bet kuriam i tikimybė $p(x_i) = \alpha_i$, kur $\alpha_i < 1$, ir $\sum_{i=1}^n p(x_i) = 1$.

Kiekvieno tipo grupei reikalingas savas, specifinis, kokybės atributų reikšmių agregavimo metodas.

Agregavimo metodai

Panagrinėsime, kokius agregavimo metodus tikslinga naudoti kiekvienu iš ankstesniame skyrelyje aptartu atveju.

Kadangi pirmuoju atveju visi agreguojami atributai yra ortogonalūs, tai akivaizdu, kad šiuo atveju agregavimo rezultatas neturėtų viršyti mažiausios iš agreguojamų reikšmių ir todėl pirmojo tipo grupių reikšmės turi būti agreguojamos panaudojant atitinkamą t -normą. Priminsime, kad mūsų taksonomijoje bet kurio agreguojamo atributo x reikšmė yra traktuojama kaip tikimybė $q(x)$, kad atributu x aprašomos kalbos savybės $L(x)$ pakaks atitinkamai bet kurios teoriškai įmanomos informacinės sistemos savybei s_x specifikuoti. Todėl pirmuoju atveju reikšmių agregavimą galima (ir

tikslinga) traktuoti kaip tikimybių sandaugą

$$q(x) = \prod_{i=1}^n q(x_i).$$

Pažymėtina, kad tikimybių sandauga tenkina t -normos savybes.

Antruoju atveju grupei priklauso bent vienas atributas, kuriuo aprašomos savybės prireiks specifikuojant tiksliai kai kurias informacines sistemas. Tai reiškia, kad, nepriklausomai nuo tokios savybės išsamumo, jos visiškai pakaks specifikuojant tokias sistemas, kurias specifikuojant jos neprisireikia. Kitaip tariant, šiuo atveju turi būti panaudota atitinkama svorinė t -norma, t.y. atributams, kuriais aprašomų savybių reikia specifikuojant tik kai kurias informacines sistemas įvedamas koeficientas $1 - p(x)$. Taigi, atitinkama svorinė t -norma yra aprašoma formule

$$q(x) = \prod_{i=1}^n \left(1 - p(x_i)(1 - q(x_i))\right).$$

Trečiuoju atveju agreguojamais atributais aprašomos kalbos savybės papildo viena kitą. Tai reiškia, kad šiuo atveju yra reikalinga atitinkama t -konorma. Be to, konstruojant agregavimo operatorių reikia atsižvelgti į tai, kad papildomų kalbos galimybių prisireikia tik tuomet, kuomet nepakanka bazinėmis savybėmis nusakytų jos galimybių. Taigi, šiuo atveju galima pasinaudoti vadinamąja papildymų ir atmetimų formule:

$$q(x) = \sum_{m=1}^n (-1)^{m-1} \sum_{1 \leq i < \dots < m \leq n} q(x_i) \cdot \dots \cdot q(x_m).$$

Nesunku parodyti, kad ši formulė tenkina t -konormos savybes.

Ketvirtuoju atveju agreguojamais atributais aprašomos kalbos savybės nebūtinai yra išsamios. Jų išsamumas nusakomas atitinkamais svoriniais koeficientais ir, atsižvelgiant į tai, kad tos savybės yra alternatyvios, šiuo atveju turi būti naudojama atitinkama svorinė t -konorma. Konstruojant konkretų agregavimo operatorių, dar reikia atsižvelgti į tai, kad kalbos savybės gali tam tikru mastu dubliuoti viena kitą. Taigi, paprasčiausias agregavimo operatorius šiuo atveju aprašomas formule:

$$q(x) = \sum_{i=1}^n q(x_i)p(x_i) - \prod_{i=1}^n q(x_i)p(x_i).$$

Išvados

Tikimybinė IS specifkavimo kalbų kokybės atributų traktuotė įgalina sukonstruoti tokią taksonominę tų atributų hierarchiją ir parinkti tokius agregavimo metodus, kad agreguojant atributų reikšmes pagal taksonomines grupes ir pereinant iš vieno hierarchijos lygmens į kitą, būtų garantuojama, jog pirminių matavimų prasmė nebus iškreipta.

Literatūra

1. J. Gasperovič, A. Čaplinskas, Informacinių sistemų specifikavimo kalbų kokybės vertinimo problema, kn.: *Informacinės technologijos 2003, Konferencijos Pranešimų Medžiaga*, Technologija, Kaunas (2003), pp. VI-1–VI-10.
2. J. Gasperovič, A. Čaplinskas, Informacinių sistemų specifikavimo kalbų lyginamosios analizės metodai, *Informacijos mokslai*, **26** (2003).
3. A. Čaplinskas, J. Gasperovic, A taxonomy of characteristics to evaluate specification languages, in: J. Barzdins (Ed.), *Computer Science and Information Technologies*, vol. 672, University of Latvia (2004), pp. 321–336.
4. A. Čaplinskas, A. Lupeikiene, O. Vasilecas, A framework to analyse and evaluate Information Systems specification languages, *Lecture Notes in Computer Science*, **2435**, 248–262.
5. M. Detyniecki, *Mathematical Aggregation Operators and their Application to Video Querying*, Phd. thesis, Universite Pierre & Marie Curie (2000).
6. M. Grabisch, S.A. Orlovski, R.R. Yager, Fuzzy aggregation of numerical preferences, in: R. Slowinski (Ed.), *Handbook of Fuzzy Sets Series*, Kluwer Academic Press (1998), pp. 31–68.
7. ISO/IEC 9126, Information Technology – Software Product Evaluation – Quality Characteristics and Guidelines for their Use, First edition, (1991-12-15), reference number ISO/IEC 9126: 1991(E).
8. C. Moraga, Neuro-fuzzy modeling between minimum and maximum, in: *Proc. of the Workshop on Computational Intelligence, Theory and Applications*, Yugoslavia (2001).

SUMMARY

J. Gasperovič, A. Čaplinskas. Specific in aggregation of characteristics describing internal quality of IS specification language

The paper discusses how to construct taxonomy of characteristics describing internal quality of IS specification language and aggregation techniques that ensure proper aggregation of measurements through the whole taxonomy.

Keywords: IS specification language, internal quality, taxonomy, aggregation method.