



Institute of Computer Science
Academy of Sciences of the Czech Republic

Tvorba ontologie huminových látek

Pavel Tyl, Martin Řimnáč, Roman Špánek, Július Štuller,
Zdeňka Linková, Josef Kozler, Barbora Antošová, Vojtěch
Váňa

Technical report No. 1098



Institute of Computer Science
Academy of Sciences of the Czech Republic

Tvorba ontologie huminových látek

Pavel Tyl, Martin Řimnáč, Roman Špánek, Július Štuller,
Zdeňka Linková, Josef Kozler, Barbora Antošová, Vojtěch
Váňa¹

Technical report No. 1098

Abstrakt:

Znalosti, které pomáhají popsat, třídit a najít potřebná data, bývají mnohdy nashromážděny mnohaletou zkušeností. Zkušenost může odejít s pracovníkem, který je jejím nositelem. Tuto situaci může řešit použití ontologií jako jednoho z nástrojů znalostního inženýrství, obecněji v rámci metod reprezentace znalostí.

Ontologie huminových látek navržená na základě analýzy existujících ontologií podobného typu umožní zachovat, zpřístupnit a opětovně použít získané zkušenosti, například pro tvorbu funkčně bohatších, flexibilních a především snadno integrovatelných systémů a aplikací.

Práce v úvodních částech předkládá výběrový přehled formalismů pro zachycení datové struktury a procesů v oblasti huminových látek s důrazem na použití ontologií a jejich možností. Především případové studie aplikací ontologií pro evidenci a organizaci dat o huminových látkách, jejich měření a pokusech, u kterých zatím není využívání ontologií běžné, ukazuje potenciál ontologií a zároveň ilustruje parametry univerzálně použitelného modelu ontologie pro tuto oblast.²

Keywords:

ontologie, huminové látky

¹Instituce:

Technická univerzita v Liberci, Fakulta mechatroniky, informatiky a mezioborových studií, Studentská 2, Liberec, Česká republika

Ústav informatiky AV ČR, v. v. i., Pod Vodárenskou věží 2, Praha 8, Česká republika

Výzkumný ústav anorganické chemie, Revoluční 84, Ústí nad Labem, Česká republika

²Poděkování:

Tento projekt je realizován za finanční podpory prostředků státního rozpočtu České republiky prostřednictvím projektu „Pokročilé sanační technologie a procesy“ č. 1M0554 programu Výzkumná centra PP2-DP01 Ministerstva školství, mládeže a tělovýchovy.

Tvorba ontologie huminových látek

Pavel Tyl^{1,2}, Martin Řimnáč¹, Roman Špánek^{1,2}, Július Štuller^{1,2},
Zdeňka Linková¹, Josef Kozler^{2,3}, Barbora Antošová^{2,3}, Vojtěch Váňa^{2*}

¹ Ústav informatiky AV ČR, v. v. i.,
Pod Vodárenskou věží 2, Praha 8, Česká republika

² Technická univerzita v Liberci
Fakulta mechatroniky, informatiky a mezioborových studií
Studentská 2, Liberec, Česká republika

³ Výzkumný ústav anorganické chemie
Revoluční 84, Ústí nad Labem, Česká republika

Abstrakt Znalosti, které pomáhají popsat, třídit a najít potřebná data, bývají mnohdy nashromážděny mnohaletou zkušeností. Zkušenost může odejít s pracovníkem, který je jejím nositelem. Tuto situaci může řešit použití ontologií jako jednoho z nástrojů znalostního inženýrství, obecněji v rámci metod reprezentace znalostí.

Ontologie huminových látek navržená na základě analýzy existujících ontologií podobného typu umožní zachovat, zpřístupnit a opětovně použít získané zkušenosti, například pro tvorbu funkčně bohatších, flexibilních a především snadno integrovatelných systémů a aplikací.

Práce v úvodních částech předkládá výběrový přehled formalismů pro zachycení datové struktury a procesů v oblasti huminových látek s důrazem na použití ontologií a jejich možností. Především případové studie aplikací ontologií pro evidenci a organizaci dat o huminových látkách, jejich měření a pokusech, u kterých zatím není využívání ontologií běžné, ukazuje potenciál ontologií a zároveň ilustruje parametry univerzálně použitelného modelu ontologie pro tuto oblast.

Klíčová slova: ontologie, huminové látky

1 Motivace

Obecně je známo, že huminové kyseliny, v obecnější rovině huminové látky, mohou značnou měrou ovlivnit růst různých plodin. Huminové látky (dále jen HS) je proto možné také použít i v oblasti fytořemediace kontaminovaných půd. V rámci řešení projektu ARTEC – 1M0554 byl zkoumán vliv použití HS pro fytořemediaci půdy v závislosti na použité plodině. Je sledován vliv interakce HS s rostlinnými substráty na růst vybraných plodin a na dynamiku řemediace, tj.

* Tento projekt je realizován za finanční podpory prostředků státního rozpočtu České republiky prostřednictvím projektu „Pokročilé sanační technologie a procesy“ č. 1M0554 programu Výzkumná centra PP2-DP01 Ministerstva školství, mládeže a tělovýchovy.

intenzitu zachycování kontaminantů plodinou z půdy. Je reálný předpoklad, že HS díky svým stimulačním účinkům na růst rostlin výrazněji ovlivní nárůst jejich kořenového systému a tím i intenzitu přechodu kontaminantů do rostliny, která bude po sklizni využita jako biopalivo.

Výše nastíněné aktivity jsou jen jednou ze součástí záběru Výzkumného centra ARTEC, které se zabývá sanačními technologiemi. V současné době dochází k aplikování získaných poznatků o jednotlivých technologiích a jejich nasazování na reálných úlohách.

Článek popisuje pilotní projekt výzkumného centra ARTEC, jehož úkolem byl návrh a vytvoření ontologie pro problematiku výroby huminových látek a popis pokusů s huminovými látkami včetně jejich výsledků.

Úvod do problematiky huminových látek je podrobně popsán v části 2. Část 3 vysvětluje hlavní důvody, proč vytvářet ontologii. V části 4 jsou popsány základní prvky ontologií, metodologie tvorby ontologií potom v části 5, výčet důležitých pojmů, definice konceptů (a jejich hierarchie), definice vlastností a omezení v částech 6, 7, 8 a 9. Vytváření instancí je popsáno v části 10. Běžné chyby a doporučení pro tvorbu ontologie jsou detailně rozebrány v části 11, shrnutí práce je v závěrečné části 12.

2 Huminové látky

Huminové látky se nacházejí v přírodních surovinách. Jejich zdrojem jsou především kaustobiolity, tj. rašelina, lignity a mladá hnědá uhlí, tzv. *oxihumolity*. V České republice jsou jedny z nejkvalitnějších oxihumolitů na světě. Oxihumolity se především zpracovávají na alkalické humáty, tj. humát sodný a humát draselný, v omezené míře pak i na jiné deriváty HS – čisté huminové kyseliny, humáty kovů a speciálně upravené oxihumolity.

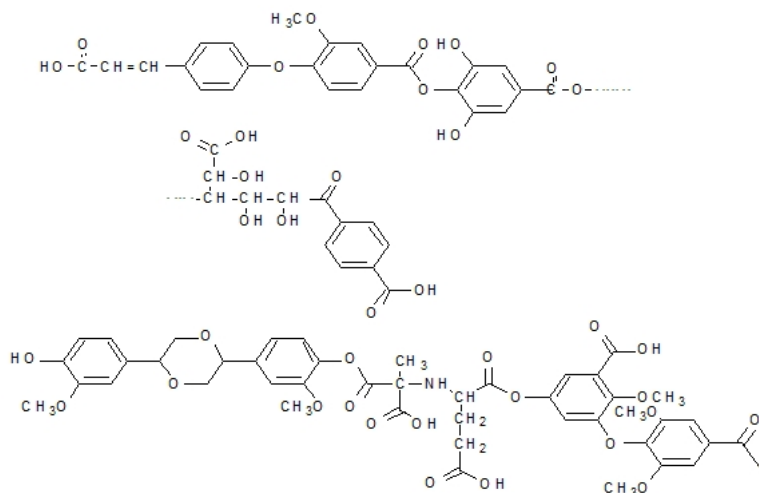
2.1 Vlastnosti huminových látek

Huminové kyseliny (HA) jsou přírodní vysokomolekulární látky s relativní molekulovou hmotností M_w 2 000 až 200 000. Jsou považovány za otevřený systém heteropolykondenzátů, polyelektrolytů aromatické a heterocyklické struktury.

Chemická struktura přírodních HA je velmi složitá. Strukturní vzorce, resp. části strukturních vzorců jsou konstruovány na základě kvantitativního stanovení základních stavebních jednotek a funkčních skupin. Pro představu uvádíme jednu z možných struktur HA, jak ji publikoval [3] (viz obr. 1).

Základními stavebními jednotkami HA jsou především fenoly a benzenkarboxylové kyseliny vázané navzájem různým způsobem a různě kondenzované s aminokyselinami. Přítomnost karboxylových ($-\text{COOH}$) a fenolických hydroxylových skupin ($-\text{OH}$), u nichž je vodík snadno nahraditelný jiným kationtem nebo radikálem, je příčinou jejich vysoké schopnosti vázat kovové prvky z kapalných (např. odpadní vody) nebo pevných (např. půda) systémů.

Přítomnost chinolinových skupin v HA vysvětluje jejich schopnost vázat aminokyseliny a umožňovat výměnu volných elektronů mezi dusíkem aminokyselin



Obrázek 1. Jedna z možných struktur HA [3].

a chinolinovými skupinami. Tyto jejich vlastnosti se vůči danému systému, ve kterém působí, projevují jako zdroj potenciální biologické aktivity, vazebných a sorpčních schopností. To se projevuje tím, že jejich přítomnost v daných systémech ovlivňuje, mnohdy významně, chemické a biochemické procesy. Díky této schopnosti mají HS vysokou perspektivu širokého uplatnění v různých oblastech technické praxe (chemický a keramický průmysl, agrochemie, veterinární a humánní medicína, stavebnictví).

V biochemických procesech se uplatní HS ve fyziologii rostlinných buněk. Výše determinovaná reaktivita HS je teoreticky srovnatelná s předpokládanou reaktivitou povrchu jednobuněčných organismů, obecně buněčných povrchů. Lze předpokládat, že takto podmíněná interakce HS a buňky je fyziologicky působícím signálem (faktorem), modelujícím buněčný fenotyp, tedy jednou z obecných podstat biologického účinku HS.

Význam HS v biochemických procesech neustále vzrůstá v přímé souvislosti s nalézáním nových možností jejich aplikací v praxi. Výčet jejich možných účinků ve fyziologii je rozsáhlý. Účinek HS v daných systémech je odvislý od jejich složení, respektive způsobu zpracování a koncentraci. Na růst mikroorganismů mohou mít jak stimulační, tak i retardační účinky.

Vysoká chemická a biologická aktivita HS se mj. projevuje tím, že příznivě ovlivňují všechny fyziologické procesy rostlin. Práce rostlinných fyziologů v 60. letech prokázaly, že na buněčném základě ovlivňují asimilaci, dýchání, energetický systém, vývojový a růstový cyklus rostlin.

2.2 Aplikační možnosti huminových látek v technické praxi

Vybrané průmyslové aplikace v chemických procesech – HA jsou svojí strukturou blízké průmyslově vyráběným ionexům, proto je možné používat oxyhumolity a deriváty HA k zachycování toxických kovů (Pb, Hg, Cd, Cu, Zn, Ni, Cr), respektive jejich iontů z odpadních vod. Stejně jako ionexy se dají použít do kolon a regenerovat, nebo se používají ve formě přísad do odpadní vody s následným oddělením sedimentací nebo filtrací. Nejvhodnější použití těchto látek je tam, kde se koncentrace toxických kovů pohybuje v rozmezí 1 až 10 mg/l. Zbytková koncentrace kovů v odpadních vodách je zpravidla nižší než 0,3 mg/l. Jsou ekologicky nezávadné a vazebná síla vůči kovům je srovnatelná s průmyslovými ionexy.

V papírenském průmyslu se uplatňují HA, resp. jejich sodná forma, humát sodný (dále jen HNa), který se používá jako barvivo papíru. Jeho přítomnost zároveň zlepšuje retenci papíroviny a snižuje průnik toxických látek do podsítných vod. Barvení je bezproblémové jak u technologie, kdy je barva přidávána přímo do hmoty rozvlákněného papíru, tak u technologie, kdy se barva na povrch papíru nanáší na klížícím lisu.

Ve stavebnictví se HS uplatňují jako účinná složka plastifikátorů betonových směsí. Plastifikátory s obsahem HA řeší základní problém transportbetonu, tj. zablokování hydratace cementu a zachování konzistence betonové směsi během její dopravy bez negativního vlivu na rychlost nárůstu počátečních pevností betonu. Plastifikátory na bázi HA jsou ve své třídě kvalitativně srovnatelné se zahraničními výrobky při nižší ceně.

Půdní aplikace – HS mají pozitivní vliv na růst a vývoj u mnoha zemědělských plodin a také u lesních porostů. Zlepšují tvorbu kořenového systému, zvyšují nárůst biomasy, zlepšují kvalitu produkce a jejich přítomnost v půdě zabráňuje nebo velmi omezuje přechod toxických prvků do zemědělských plodin. HS se používají samostatně nebo v kombinaci s kapalnými a tuhými hnojivy.

Oxyhumolity lze přidávat do kontaminovaných půd, výhodně pak v kombinaci s vápněním. HA obsažené v oxyhumolitech zachycují těžké kovy a organické polutanty. HA v oxyhumolitech jsou kvalitnější než HA v rašelině.

Roztoky HS se využívají k postřiku na list v kombinaci s výživou rostlin. HA působí jako stimulatory místa a inhibitory vzniku plísní. Velmi perspektivní je možnost využití HS k remediace půd po průmyslových a důlních zátěžích. Zde se především uplatní různé formy upravených oxihumolitů. Jedna z forem jejich možných aplikací jsou substráty pro plodiny za použití.

Humánní a veterinární medicína – výzkum VÚAnCh a specializovaných pracovišť ukazuje, že se HA mohou uplatňovat v léčivé formě, a to jak pro aplikaci u lidí, tak i u zvířat. Jejich účinnost spočívá v jejich schopnosti vázat rizikové chemické prvky, ovlivňovat fyziologické procesy a pohlcovat UV záření.

3 Proč vytvářet ontologii?

V posledních letech se vývoj ontologií – explicitní formální specifikace termínů určité domény a vztahů mezi nimi – přesouvá z oblasti umělé inteligence do

oblastí zájmu expertů v různých vědních oborech. Ontologie se rovněž stávají běžnou součástí Internetu. Ontologie na Webu zahrnují rozsáhlé taxonomie, ale mohou zahrnovat například i kategorizaci produktů s jejich vlastnostmi. Mnoho vědních disciplín vyvíjí standardizované ontologie, které doménoví experti mohou použít ke sdílení a anotaci dat jejich oblasti zájmu. Například v oboru biochemie vznikl velký standardizovaný strukturovaný slovník a ontologie chemických entit biologického zájmu – ChEBI [I4]. Konsorcium W3C [I11] navrhlo RDF – Resource Description Framework („rozhraní pro popis zdrojů“) a později jazyky OWL a OWL 2.0 [I6], jazyky pro popis znalostí o objektech (entitách) uvedených na webových stránkách tak, aby s nimi mohli pracovat i softwaroví agenti hledající určité informace.

Ontologie tedy definuje určitý slovník pro skupiny uživatelů (lidí nebo softwarových agentů), kteří potřebují sdílet data, informace a znalosti o určité doméně. Zahrnuje počítačově zpracovatelné definice pojmů dané domény a vztahů mezi těmito pojmy.

Proč konkrétně vytvářet ontologii? Následují některé z možných důvodů:

- *Ke sdílení dat, informací a znalostí mezi lidmi a softwarovými agenty* – jeden z nejčastějších cílů vývoje ontologií, který dovoluje softwarovým agentům shrnout extrahované informace z více překrývajících se zdrojů a odpovědět tak na uživatelem formulované dotazy.
- *Ke znovupoužití doménových znalostí* – budeme-li chtít vytvořit novou ontologii, lze pro to použít ty, které již existují a integrovat je nebo naopak podle své potřeby rozšířit či zúžit.
- *K vytvoření aktuálního obrazu znalostí o určité doméně* – jednoznačná specifikace doménových znalostí je důležitá zejména pro nové uživatele, kteří potřebují vědět, co který termín v doméně znamená. Jsou-li poznatky zapsány v ontologii, je snadné je najít a případně změnit. Měnit a hledat všechny softwarové agenty s pevně zapsanými poznatky (ve formě vnitřní logiky každého konkrétního agenta) by bylo mnohem složitější.
- *K analýze doménových znalostí* – tu je možné provést, je-li dostupný přesný význam jednotlivých pojmů ontologie. což může být následně využito pro rozšíření a znovupoužití existujících ontologií.
- *K oddělení znalostí doménových a provozních* – další běžné využití ontologií. Je například možné popsat konfiguraci nějakého produktu z jeho komponent podle požadované specifikace a vytvořit program, který provede konfiguraci nezávisle na konkrétním produktu a komponentách samotných. Jedním programem jsme pak například schopni konfigurovat „cokoli“ od osobního počítače po server.

Doménová ontologie přesto nebývá cílem sama o sobě. Ontologie se obvykle vytváří proto, aby ji mohli používat jiné programy. Doménově nezávislé aplikace nebo softwaroví agenti mohou používat ontologie a znalostní báze jako část svých vstupů. V tomto kontextu vytvořená ontologie huminových látek slouží jak pro interní spolupráci v rámci řešitelského týmu, tak pro prezentaci získaných poznatků a znalostí v oblasti huminových látek třetím stranám.

Je téměř nemožné pokrýt všechny požadavky, které může vývojář i uživatel rozsáhlejší ontologie potřebovat. Stejně tak neexistuje „jediný správný“ postup pro tvorbu ontologie. Tento příspěvek se snaží naznačit jeden z možných způsobů tvorby konkrétní ontologie.

4 Co je v ontologii?

Definice ontologie je více. Pro účel tohoto článku bude ontologie explicitní formální popis pojmů dané domény, v našem konkrétním případě tedy domény huminových látek. Ontologie bude obsahovat *objekty* (zvané *koncepty*, někdy též *třídy*), jejich *vlastnosti* (zvané *role*, někdy též *sloty*) a daná *omezení* kladená na *objekty* a *vlastnosti* (*omezení rolí*, někdy též *aspekty*). Ontologie spolu s *instancemi konceptů* a *rolí* tvoří *znalostní bázi*. Hranice přechodu mezi ontologií a znalostní bází však bývá mnohdy vágní.

Koncepty jsou základem všech ontologií. Koncepty popisují základní pojmy domény. Například koncept *Surovina* reprezentuje obecnou surovinu. Konkrétní existující suroviny jsou pak **instancemi** tohoto konceptu. Koncept může mít *podkoncepty*, které jsou specifitější než koncept samotný. Například *Pokus* má podkoncepty *Parcelový-pokus* a *Nádobový-pokus*. **Hierarchie** konceptů (vztah podkoncept–nadkoncept) je tranzitivní.

Role popisují vlastnosti konceptů a instancí konceptů. Každá *Surovina* má například určité *majoritní-prvky* a *sledované-prvky*. Vlastnost *majoritní-prvky* může nabývat několika různých hodnot (*Al*, *Ca*, *Cl*, *Fe*, ...). Vlastnost *sledované-prvky* může nabývat dalších možných hodnot (*As*, *Cd*, *Co*, *Cr*, ...). Na úrovni konceptu tedy můžeme říci, že instance konceptu *Surovina* budou mít role *majoritní-prvky*, *sledované-prvky*, ...

Všechny **instance** konceptu *Surovina* mají roli *majoritní-prvky*, která udává hodnotu koncentrace jednotlivých prvků (instancí konceptů *Prvky*) v *Surovině* (viz obr. 2).

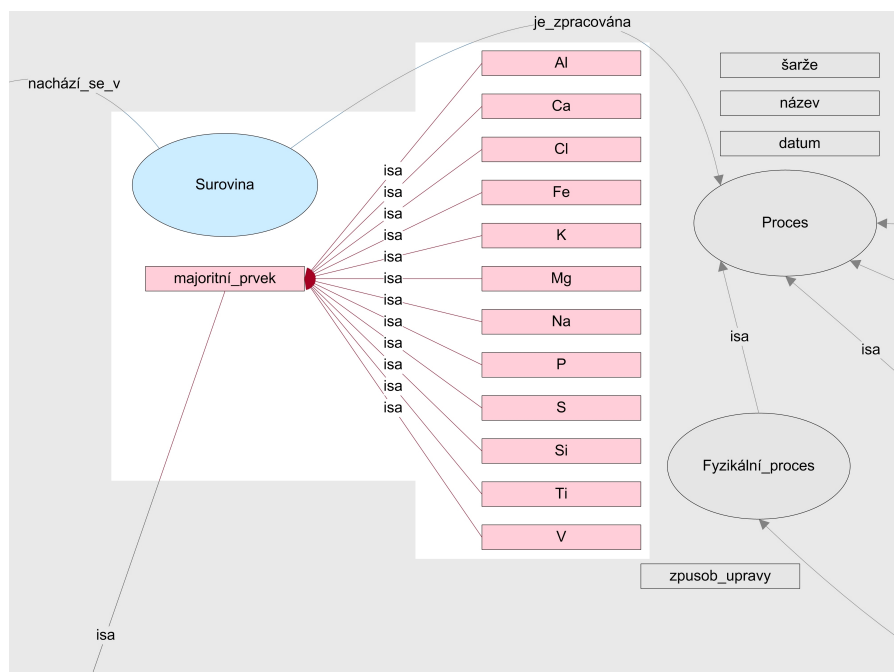
Prakticky lze říci, že tvorba ontologie zahrnuje:

- *definici konceptů* ontologie,
- *uspořádání konceptů* do taxonomické (podkoncept–nadkoncept) hierarchie,
- *definici rolí* a *popis povolených hodnot a omezení těchto rolí*,
- někdy také *doplnění instancí* a *hodnot pro role* (viz obr. 3).

5 Metodologie tvorby ontologie

Základní pravidla, která je dobré dodržet a uvědomit si při vytváření ontologie jsou tato:

1. Neexistuje „jedna správná“ metoda pro vývoj ontologie – při modelování domény zájmu jsou většinou *možné alternativy*. „Nejlepší“ řešení často závisí na aplikacích využívajících danou ontologii a dalších předvídaných rozšířeních.



Obrázek 2. Koncept *Surovina* a role *majoritní_prvek*.

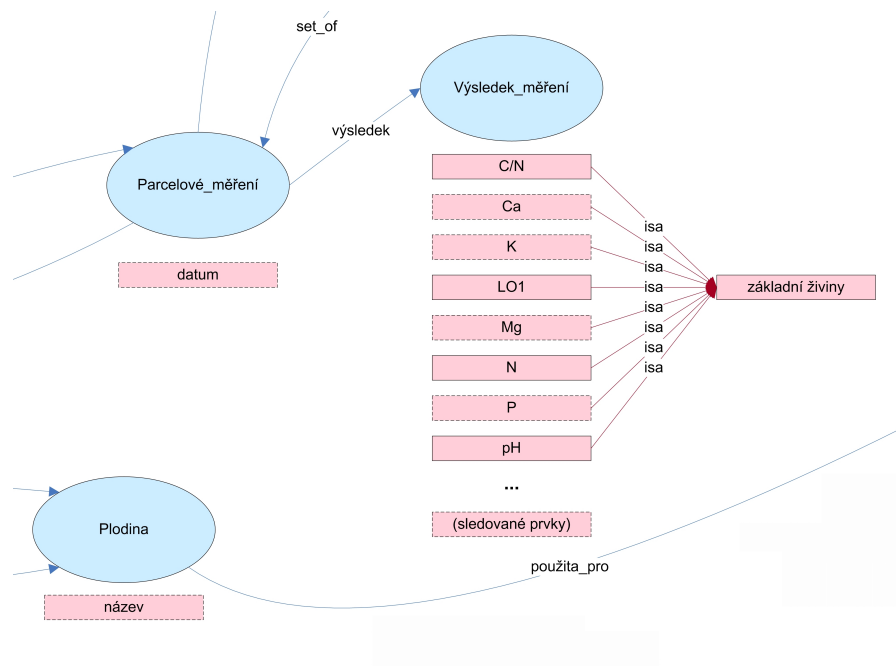
Modrou barvou jsou ve zvýrazněné oblasti znázorněny koncepty, červenou datotypové role a instance. Šipky znázorňují objektové role, šipky se slovem *is-a* představují role podkoncept–nadkoncept, příp. podvlastnost–nadvlastnost

2. Vývoj ontologie je nutně *iterativní proces* – začínáme zběžným průchodem ontologií, opravujeme a vylepšujeme ji, doplňujeme detaily (během toho vyhodnocujeme klady a zápory svých rozhodnutí).
3. Pojmy ontologie by se měli co nejvíce blížit realitě popisované domény. (*Koncepty* budou pravděpodobně znázorněny *podstatnými jmény*, *role* mezi nimi *slovesy* a společně budou vytvářet jednoduché věty popisující doménu zájmu.)

Rozhodnutí, k čemu a jak bude sloužit naše ontologie, popřípadě jak detailní nebo obecná ontologie bude, hodně ovlivní její návrh. U možných alternativ musíme určit:

- která lépe splní náš úkol,
- která bude intuitivnější,
- která půjde lépe rozšířit,
- která půjde lépe udržovat.

Poté, co definujeme základní verzi ontologie, můžeme ji vyhodnocovat a opravovat s pomocí specializovaných programů či expertů v oboru. *Revize a iterace* jsou v *životním cyklu ontologie* velice důležité.



Obrázek 3. Některé koncepty, jejich role a relace mezi nimi v doméně huminových látek.

5.1 Určení oblasti a rozsahu ontologie

Při určení oblasti a rozsahu ontologie je potřeba zodpovědět několik základních otázek:

- Jakou doménu bude ontologie pokrývat?
- K čemu bude ontologie využita?
- Na jaké typy otázek by data, informace a znalosti popisované ontologií měly poskytnout odpovědi?
- Kdo bude ontologii používat a spravovat?

Odpovědi na tyto otázky se mohou během procesu návrhu ontologie měnit, ale ve kterýkoli okamžik pomáhají s vymezením rozsahu ontologie.

Doménou námi navržené ontologie jsou huminové látky, které se v rámci tohoto článku používají jako aditiva k rostlinným substrátům, ale v obecné rovině mohou být používány pro specifikované aplikace v daných chemických nebo biochemických procesech. Ontologii bude možné využít:

1. pro kvalitativní posuzování a potažmo pro výběr ložisek primárních uhlíkatých surovin huminových kyselin, v obecné rovině huminových látek (oxyhumolity, lignit, rašelina),

2. pro postup zpracování primární uhlíkaté suroviny na produkty (mechanicky upravený oxyhumolit, alkalické humáty, čisté huminové kyseliny) o požadovaných kvalitativních parametrech huminových kyselin v nich obsažených (obsah HA, spektrum relativních molekulových hmotností, úroveň jejich biologické aktivity, obsah funkčních skupin, které mají zásadní vliv na jejich chemickou a biologickou aktivitou [–COOH a fenolické –OH skupiny], obsah popela),
3. pro posuzování vhodnosti aplikace huminové látky o daných parametrech kvality k procesům, s definovanými podmínkami jejich uspořádání nebo naopak, jaké kvalitativní parametry musí mít huminová látka, aby vyhovovala požadavkům námi definovaného procesu.

Pokud by námi vytvářená ontologie byla použita pro zpracování textů o chemických látkách v přirozeném jazyce, bylo by jistě vhodné vložit do ontologie synonyma a informace o slovních druzích pro pojmy v ontologii. Bude-li ontologie použita pro stanovení ekonomické náročnosti, bylo by nutné zahrnout finanční náklady na jednotlivé pokusy či měření, případně dostupnost určitých surovin či chemikálií. V případě, že jazyk, ve kterém bude ontologie navržena a spravována bude odlišný od jazyka lidí, kteří tuto ontologii budou používat, měli bychom poskytnout mapování mezi těmito jazyky.

5.2 Na jaké otázky dá ontologie odpovědi?

Jedním ze způsobů, jak určit zaměření ontologie, je nastínění typu otázek, na které databáze, informační systém či znalostní báze založené na ontologii, dokáží odpovědět, tzv. otázek kompetence [1]. Tyto otázky mohou posloužit jako test toho, zda je ontologie navržena správně:

- Obsahuje ontologie dostatek informací k zodpovězení těchto typů otázek?
- Jsou odpovědi na otázky na dostatečně detailní úrovni?

Otázky kompetence slouží jako nástin, nemusí být kompletní a vyčerpávající. V ontologii huminových látek jsou otázky kompetence následující:

- Jakou energetickou rostlinu pěstovat, je-li nějaká oblast kontaminována „některým ze sledovaných prvků“? Výběr vhodných rostlin záleží vždy na konkrétním případě a podmínkách, které výběr ovlivňují, především: typ a druh kontaminantu, podmínky stanoviště (klimatické, půdní, vlhkostní apod.), použitý technologický systém (např. fyto Remediacce jako součást jiných remediálních aplikací na stanovišti), technologie pěstování jednotlivých rostlin, případná možnost jejich dalšího využití (fytoenergetika, protierozní opatření, ekologické a krajinářské funkce). V případě fytoenergetického využití připadají v úvahu vytrvalé vysokovzrůstné druhy, které sice nemají hyperakumulární efekt, ale tento je „nahrazen“ produkcí velkého množství fytomasy. Pro vlhčí stanoviště je z tohoto pohledu vhodná chřastice rákosovitá (*Phalaris arundinacea*) jako zástupce trav, z dřevin pak vrba košíkářská (*Salix viminalis*). Pro standardní stanoviště se jeví perspektivní smoloroň (mužák)

prorostlý (*Silphium perfoliatum*), šťovík Uteuša (*Rumex tianshanicus* × *Rumex patientia*), ze zástupců čeledi bobovitých rostlin např. čičorka pestrá (*Coronilla varia*).

- Jaké finanční náklady budou potřeba pro dekontaminaci v rozmezí pěti let? Výše finančních nákladů pro dekontaminaci závisí na několika faktorech – jsou to především: vybraná rostlina a náročnost její pěstební technologie včetně použití různých druhů mechanizačních prostředků, typ a způsob aplikace huminových látek, klimatické faktory a podmínky stanoviště, stupeň intenzity pěstební technologie (extenzivní, standardní, intenzivní), další využití narostlé fytomasy (energetické využití – vlastní, prodej). Na základě výše uvedených faktorů se náklady na dekontaminaci pomocí rostlin (celkem za prvních pět let) pohybují od cca 45 tis. Kč do 125 tis. Kč na hektar. Také s přihlédnutím k finančním nákladům jsou preferovány víceleté nebo vytrvalé rostliny, kde musíme oproti jednoletým rostlinám počítat se zvýšenými náklady pouze v prvním roce při zakládání porostů (u jednoletých se porost zakládá každý rok). Na druhé straně oproti výši nákladů jsou potenciální příjmy z případného prodeje produktu (v tomto případě fytomasy) a případná dotační podpora pro pěstování energetických rostlin.

Vyhodnocením těchto otázek získáme přehled, jaké možné odpovědi (ve formě příslušných dat, informací a znalostí) lze získat z databáze, informačního systému či báze znalostí vybudované nad vytvořenou ontologií – například volba nejvhodnější energetické rostliny při kontaminaci, finanční a časové náklady na dekontaminaci atd.

5.3 Zvážení použití již existujících ontologií

Vždy stojí za to využít práce, kterou někdo danému problému věnoval a zjistit, je-li možné vylepšit, popřípadě rozšířit existující zdroj pro naši studovanou doménu. Znovupoužití existující ontologie může být dokonce nutné v případě, že používáme aplikace, které aktuálně nějakou ontologii či slovník využívají. V elektronické podobě je v současné době mnoho ontologií, které lze většinou nástrojů importovat do prostředí pro vývoj ontologií, které používáme. Nezáleží často ani na formalismu, ve kterém je ontologie vyjádřena, navíc převod mezi nimi většinou není obtížný.

Knihovny použitelných ontologií na Webu jsou například databáze a ontologie Ontolingua [I5] nebo ontologie DAML [I2]. Existuje rovněž množství veřejně dostupných nekomerčních ontologií, například UNSPSC [I10]. Již vzpomínaná ontologie ChEBI [I4] se zaměřuje na klasifikaci chemických látek biologického významu. Existuje také taxonomie chemických látek, která vznikla za spolupráce HCLS (Semantic Web Health Care and Life Sciences Interest Group) a W3C [I3].

Importem všech těchto ontologií by se však naše ontologie příliš odchýlila od problematiky pokusů, měření, sledování ekonomických a dalších aspektů huminových látek. Předpokládejme proto, že aktuálně neexistuje žádná vhodná relevantní ontologie pro náš účel.

6 Výčet důležitých pojmů ontologie

Je užitečné si napsat seznam pojmů, které se budou v ontologii určitě vyskytovat. U každého pojmu je důležité si shrnout jeho vlastnosti a popis. V ontologii huminových látek patří mezi tyto pojmy například *Surovina*, *Lokalita*, *Biologická_aktivita*, *Parcela*, *Plodina*, *Substrát*, *Energetická_rostlina*, *Fyzikální_proces*, *se_měří*, *se_zpracovává*, ... Seznam pojmů by měl být úplný, vyčerpávající. Pojmy samotné se přitom mohou překrývat.

Dalšími dvěma (a většinou těmi nejdůležitějšími) kroky vývoje ontologie je definování hierarchie konceptů a jejich rolí. Není snadné tyto dva kroky dělat odděleně, většinou vytvoříme několik konceptů a popíšeme je pomocí rolí.

7 Definice konceptů a hierarchie konceptů

Existuje několik možných přístupů, jak vytvářet hierarchii konceptů – tzv. přístup *top-down*, *bottom-up* a *jejich kombinace* [2]. Žádná z metod přitom není univerzálně nejlepší nebo lepší než jiná, použití závisí na úhlu pohledu návrháře, popisované oblasti atd. V našem případě jsme využili právě kombinace obou přístupů. Výše uvedené pojmy bylo nutné někdy „obalit“ jedním obecnějším (postup *bottom-up*), jindy zase v hierarchii více specifikovat (postup *top-down*).

Ať už vybereme jakýkoli přístup, obvykle začínáme definicí pojmů. Ze seznamu „důležitých“ pojmů (viz část 6) vybereme nejdříve ty, které existují samy o sobě, až potom ty, které právě tyto nezávislé nějakým způsobem popisují. Pojmy, které existují samy o sobě, jsou v ontologii obvykle reprezentovány koncepty. Hierarchii konceptů vytváříme tak, že se ptáme, zda instance jednoho konceptu bude nutně zároveň instancí konceptu jiného (například vnořeného). Například *Cena* jakéhokoli pojmu ontologie je nutně *Ekonomickým_parametrem*.

8 Definice vlastností konceptů – rolí

Koncepty samotné neposkytnou dostatek informací k zodpovězení otázek kompetence (viz část 5.2). Pokud jsme již definovali nějaké koncepty, musíme popsat také jejich vlastnosti či vztahy mezi nimi. Ze seznamu pojmů jsme již vybrali koncepty, ostatní budou již pravděpodobně jejich rolemi. Jsou to například *se_měří*, *je_zpracována*, *nachází_se*, *leží_v_lokalitě*, ...

Pro každou roli na seznamu musíme určit, který koncept popisuje. Tyto role budou připojeny ke konceptům. Například koncept *Surovina* bude mít role *název*, *popis*, *celková_acidita* a další *fyzikálně-chemické_parametry*.

Obecně existuje několik typů vlastností, které mohou být rolemi v ontologii:

- vnitřní vlastnost – například *celková_acidita Suroviny*,
- vnější vlastnost – například *název Lokality*,
- části – pokud je objekt strukturovaný (fyzicky či abstraktně), například *relativní_přírůstek Biologické_aktivity*,

- vztahy – při spojení konkrétních instancí (daných konceptů), například *Surovina je_zpracována Procesem*.

Koncept *Surovina* tedy bude mít následující role: *název*, *celková_acidita*, *karboxylová_skupina*, *popel*, *sušina*, *voda* a další. Všechny případné podkoncepty dědí role nadkonceptu. Každý podkoncept konceptu *Proces* (v našem případě tedy *Fyzikální_proces* a *Chemický_proces*) bude mít *datum*, *název* a *šarži*. Role by měla být připojena k co nejobecnějšímu konceptu v hierarchii.

9 Definice omezení rolí

Role mohou mít různá omezení popisující například jejich povolené typy či hodnoty, počet možných hodnot (kardinalitu) a další. Například hodnotou námi navržené role *majoritní_prvky* je jeden znakový řetězec (1 až 3písmenný). Role *nachází_se* může mít vícenásobné hodnoty, jež jsou instancemi konceptu *Místo_odběru*.

Nejčastější omezení rolí jsou tato:

Kardinalita role – definuje kolik hodnot může mít role. Některé systémy rozlišují pouze mezi kardinalitou jedna a vícenásobnou, některé dovolí specifikovat minimum a maximum.

Datový typ role – vyjadřuje, jaké datové typy mohou mít hodnoty role. Nejčastějšími typy jsou řetězec, číslo, pravdivostní hodnota, výčet a instance.

Doména role – co nejobecnější koncept, který role popisuje. *Surovina* je doménou role *je_zpracována*.

Rozsah role – co nejobecnější koncept, jehož instance jsou povolenými hodnotami role. *Proces* je rozsahem role *je_zpracována*. Je-li rozsahem role *je_zpracována* jak *Proces*, tak *Chemický_proces*, použijeme pouze koncept nejvýše v hierarchii, tedy *Proces*.

10 Vytváření instancí

V některých případech posledním krokem tvorby ontologie je vytváření instancí konceptů. Definice instance vyžaduje:

1. výběr konceptu,
2. vytvoření instance konceptu,
3. naplnění hodnotami rolí.

Chceme-li tedy vytvořit instanci *Nádobového_pokusu* (jeden konkrétní pokus), musíme doplnit například následující hodnoty rolí:

- *datum*: 2008-04-05,
- *hnojení*: Před zasetím každé plodiny se provede rozbor substrátů, který se bude aplikovat v nádobách a dle výsledků se stanoví dávky živin (biogenních prvků) pro pěstované plodiny. Ty se pak aplikují formou roztoků při zálivce,
- *použité_půdy* (*substráty*):
 - Lokalita *Praha – Ruzyně* (označení *R*),

- Lokalita *Milevo* (označení *M*),
- *označení_nádob*:
 - *R* (varianty 1–10),
 - *M* (varianty 11–20),
- *počet_opakování*: 2 (a, b),
- *schéma_pokusu*:

Číslo varianty	Substrát	Označení nádob	Plodina	Hnojení [g/nád.]	Uhlí [g/nád.]
1	R	1R	–	–	–
2	R	2R	–	ano	–
3	R	3R	–	ano	60
4	R	4R	–	ano	120
5	R	5R	–	ano	240
6	R	6R	ano	–	–
7	R	7R	ano	ano	–
8	R	8R	ano	ano	60
9	R	9R	ano	ano	120
10	R	10R	ano	ano	240
11	M	11M	–	–	–
12	M	12M	–	ano	–
13	M	13M	–	ano	60
14	M	14M	–	ano	120
15	M	15M	–	ano	240
16	M	16M	ano	–	–
17	M	17M	ano	ano	–
18	M	18M	ano	ano	60
19	M	19M	ano	ano	120
20	M	20M	ano	ano	240

- *uhlí*: pracovní označení *JV21* (lom Družba, Sokolov),
- *zálivka*: na 50 % maximální vodní kapacity,
- *zajištění_pokusu*: Pokus byl umístěn tak, aby při vysokých srážkových úhrnech nedošlo ke ztrátám proteklého roztoku z misek pod vegetačními nádobami,
- *použitá_plodina*: Každý půdní typ má část nádob bez rostlin a část ovlivněnou pěstováním rostlin (instance konceptu *Plodina*)

	2005	2006	2007	2008
1. plodina	–	pšenice jarní	pšenice jarní	oves
Následná plodina	kukuřice setá	peluška	kukuřice setá	–

- *použité analytické metody*:
 - *Půdní vzorky*

pH/KCl
 Výluh Mehlich II. – obsah P, K, Ca, Mg [mg/kg]
 Obsah humusu [%]
 Celková organická hmota v sušině [%]
 $HS_A + HS_B$ v sušině [%]
 - *Rostlinné vzorky*

Stanovení koncentrace majoritních prvků (N, P, K, Ca, Mg) v sušině [%],
- *má_výsledek*: Výsl_ER_2008-04-05_001 (instance konceptu *Výsledek_pokusu*).

10.1 Vysvětlení k vytvořeným instancím

2005

V průběhu 1. roku nádobového pokusu nebyly zaznamenány změny v obsahu huminových látek v zemině v důsledku přeměny uhlí. Z výsledků stanovení celkové organické hmoty vyplývá, že v rámci chyby, zvýšení odpovídá přídávkům. Výsledky analýz a hmotnost sušiny sklizené kukuřice nekorespondují se stupňovanými dávkami uhlí. Pokles koncentrace prvků v nádobách s oxyhumolitickým uhlím odpovídá zředovacímu efektu nárůstu hmotnosti sušiny.

2006

V roce 2006 bylo při sklizni plodin zjištěno, že vysoké dávky uhlí snižují hmotnost sklizených rostlin. Při rozboru rostlin pšenice, která byla pěstována na obou půdních typech, byly nalezeny vyšší koncentrace N, P a Ca u všech nádob s aplikací uhlí bez ohledu na výši dávky. U pelušky byl nárůst koncentrace těchto prvků pozorován jen na substrátu z Ruzyně. Rozborem substrátů po ukončení pokusu bylo zjištěno, že stupňované dávky uhlí zvyšovaly obsah stanoveného humusu a snižovaly hodnotu pH substrátu.

2007

Výsledky hmotnosti sušiny sklizené pšenice v 1. pololetí 2007 nebyly ovlivněny dávkou uhlí v nádobách. Následnou plodinou byla kukuřice setá. Hmotnost sušiny sklizené kukuřice byla zvýšena stupňovanými dávkami oxyhumolitického uhlí. Rozbor substrátů na podzim 2007 ukázal, že obsah celkové organické hmoty v sušině odpovídá přídávkům oxyhumolitického uhlí.

Založený nádobový pokus s oxyhumolitickým uhlím splňuje předpoklady úspěšného pokračování. Výsledky nejsou úplné a budou průběžně doplňovány.

11 Definice konceptů a hierarchie konceptů – běžné chyby, doporučení

V této části budou uvedeny některé běžné chyby při definici konceptů a jejich hierarchie, dále pak doporučení, která pomohou k jejich řešení.

V rámci navržené ontologie může hierarchie záviset na možných využitích dané ontologie, úrovni detailů nutných pro aplikační zpracování, ale také na požadavcích kompatibility s jinými existujícími modely. Existuje však určitá metodika, kterou je vhodné se řídit. Potom, co jsme definovali požadované koncepty, je vhodné hierarchii zpětně zkontrolovat.

11.1 Je hierarchie správná?

IS-A vztah – Hierarchie konceptů reprezentuje IS-A (KIND-OF) vztah: koncept *Chemický proces* je podkonceptem *Proces*, pokud každý konkrétní chemický proces je procesem. Podkoncept konceptu představuje pojem, který určitým způsobem reprezentuje nadkoncept. Chybou by například bylo navrhnout koncept *Parcelové měření* jako podkoncept *Parcelového pokusu*. Přestože se parcelové

pokusy skládají ze sady (SET-OF) parcelových měření, nemůžeme říci, že každé konkrétní *Parcelové_měření* je *Parcelovým_pokusem*, protože by to neodpovídalo realitě.

Tranzitivita hierarchických vztahu – vztah podkonceptů je tranzitivní. Je-li *Cu Sledovaným_prvkem* a *Sledované_prvky Kvantitativními_parametry*, pak *Cu* je rovněž *Kvantitativním_parametrem*. *Cu* je zároveň přímým podkonceptem *Sledovaných_prvků*.

Vývoj hierarchie konceptů – udržovat hierarchii konceptů v konzistentním stavu může být náročné, pokud se oblast zájmu vyvíjí. Je však vždy nutné, aby ontologie věrně zobrazovala námi vybraný úsek sledované domény.

Vyhnutí se cyklům v hierarchii – V hierarchii je cyklus, pokud koncept *B* je podkonceptem konceptu *A* a zároveň koncept *A* je podkonceptem konceptu *B*. To vede logicky k tomu, že koncepty *A* a *B* jsou ekvivalentní.

Analýza sourozenců v hierarchii konceptů – Sourozenci jsou přímé podkoncepty jednoho společného konceptu. Všichni by měli být na stejné úrovni obecnosti. Například *Majoritní_prvky* a *Hg* nemohou být na stejné úrovni. *Majoritní_prvky* je mnohem obecnější pojem než *Hg*. Koncepty u vrcholu hierarchie reprezentují základní pojmy a představují většinou přímý podkoncept nějakému velmi obecnému konceptu jako *Věc (Thing)*. U nich proto není potřeba, aby byly na stejné úrovni obecnosti (nemusí to být související pojmy).

Je sourozenců „moc“ nebo „málo“? – Pro počet přímých podkonceptů není přesné pravidlo. Avšak pokud je podkoncept pouze jeden, většinou to znamená, že ontologie není kompletní nebo se jedná o chybu v modelování. Toto pravidlo lze přirovnat například k seznamu – neměl by mít pouze jednu položku. Je-li podkonceptů přibližně dvanáct nebo více, je vhodné (odpovídá-li to realitě) přidat nějakou mezikategorii. Výjimku tvoří v tomto případě listové koncepty hierarchie, které by bylo případně možné nahradit instancemi (v našem případě názvy prvků – *As*, *Cd*, *Co*, *Cr*, ...). Pokud ve skutečnosti mezikategorie neexistují, samozřejmě je nevytváříme, reálný obraz má vždy přednost.

11.2 Vícenásobná dědičnost

Většina systémů pro vytváření, údržbu a zpracování ontologií (např. ontologické editory Protégé [I7], SWOOP [I9], ...) dovoluje vícenásobnou dědičnost. To znamená, že jeden koncept může být podkonceptem několika konceptů. Je jistě možné si představit množinu procesů *Fyzikálně-chemických*, jejichž instance by byly prvky jak konceptu *Fyzikální_proces* tak konceptu *Chemický_proces*. Takto nově vzniklý koncept má dva nadkoncepty a zdědí role obou rodičů.

11.3 Kdy udělat nový koncept? (A kdy ne?)

Jedno z nejtěžších rozhodnutí během modelování ontologie je, zda vytvořit nový koncept nebo vyjádřit rozdílnost pojmů pomocí různých rolí. Bývá obtížné se zorientovat jak v hierarchii s velkým množstvím vnořených konceptů, tak v ploché hierarchii s malým množstvím konceptů, ale s příliš mnoha informacemi uloženými pomocí rolí. Mezi těmito stavby je potřeba najít rovnováhu.

Existuje několik pravidel pro vytvoření nového konceptu v hierarchii. Podkoncepty nějakého konceptu mají obvykle

- navíc vlastnosti, které jejich nadkoncept nemá nebo
- omezení odlišná od nadkonceptu nebo
- se účastní jiných vztahů než nadkoncept.

Například koncept *Chemický_proces* má roli *dobu_reakce*, která se k popisu jejího nadkonceptu *Proces* nepoužívá. Nový koncept pak v hierarchii konceptů zadefinujeme obvykle tedy, když existuje něco, co o tomto konceptu můžeme říci, ne však o jeho nadkonceptu. Prakticky: každý podkoncept by měl mít buď přidáné nové role nebo definované nové hodnoty rolí nebo přidávat omezení pro zděděné role. Neměli bychom však vytvářet podkoncepty konceptů s každým přidáním omezením, které by při odvozování měly nulový přínos. Pro tato pravidla existuje výjimka – koncepty v terminologických hierarchiích (taxonomie) nemusí mít vždy nové vlastnosti, přesto je užitečné podkoncept vytvořit, protože usnadňuje procházení a navigaci v hierarchii a umožňuje snadno najít úroveň zobecnění pro konkrétní situaci. Dalším důvodem, proč vytvořit podkoncept bez nových rolí apod. může být i případ modelování reality, kdy doménový expert běžně rozlišuje určité pojmy, ačkoli nejsme nuceni rozlišení samotné modelovat. Snažíme se však reflektovat pohled experta.

11.4 Nový koncept nebo vlastnost?

Během modelování určité oblasti se často musíme rozhodnout zda modelovat specifické odlišnosti pomocí rolí nebo jako množinu konceptů. Opět to závisí na zaměření domény ontologie.

Budeme v naší ontologii vytvářet koncept *Parcelový_pokus* a *Nádobový_pokus* nebo vytvoříme jednoduše pouze koncept *Pokus* a doplníme roli *typ_pokusu* hodnotami *Nádobový* a *Parcelový*? Jak důležitým je koncept *Nádobový_pokus* v naší ontologii? Pokud by *Pokus* byl jen okrajovým konceptem v naší ontologii a pokud to, zda je či není *Nádobový* nemá další dílčí vztahy k jiným objektům ontologie, neměli bychom vytvářet zvláštní koncept nádobových pokusů. Jelikož jsou však výsledky *Nádobového_pokusu* svými vlastnostmi úplně odlišné od *Výsledků_měření_Parcelového_pokusu* a zároveň mají vztahy k jiným dalším konceptům, je velmi vhodné oba koncepty pokusů odlišit.

Jestliže se koncepty s odlišnými hodnotami rolí stanou omezujícími pro různé role, vytvoříme nový koncept pro jejich odlišení. Jinak reprezentujeme odlišnost hodnotou role. Jestliže odlišnosti jsou v doméně důležité a uvažujeme o pojmech s odlišnými rozlišujícími hodnotami jako o jiných druzích pojmů, vytváříme nový

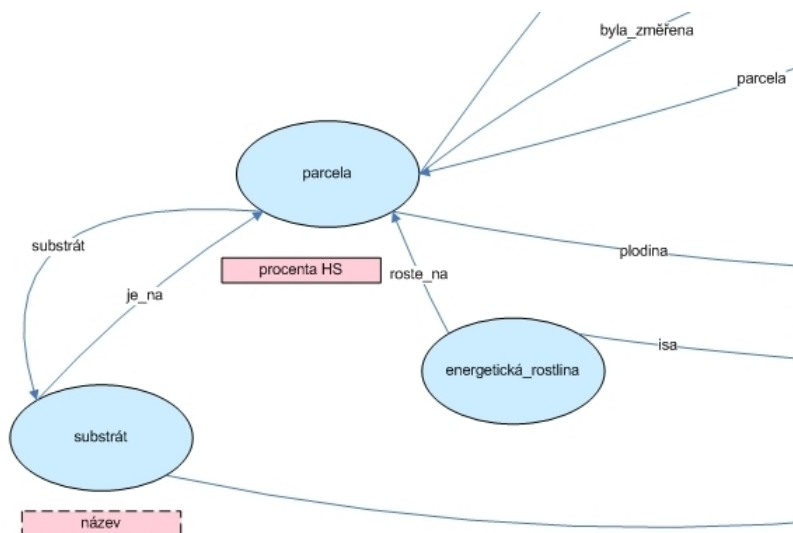
koncept. Velice vhodné je v těchto případech uvažovat o potenciálních instancích. Koncept, do něž patří konkrétní instance, by se však neměl často měnit.

11.5 Instance nebo koncept?

Rozhodnutí, reprezentujeme-li určitý pojem v ontologii konceptem nebo instancí závisí opět na potenciální aplikaci ontologie. Hranice, kde končí koncepty a začínají instance je nejnižší úroveň v hierarchii pojmů. Jaké jsou nejspecifičtější pojmy naší domény zájmu? To jsme zodpověděli v otázkách kompetence (viz část 5.2). Právě ty pojmy, které poskytují odpovědi na otázky, jsou vhodnými kandidáty na instance v naší ontologii. Instance jsou tedy nejspecifičtější pojmy v naší ontologii.

Musíme si uvědomit, že jenom koncepty mohou být uspořádány v hierarchii, instance ne, pojem subinstance se ve znalostních systémech standardně nepoužívá. Jestliže je tedy mezi objekty přirozená hierarchie, měli bychom definovat tyto objekty jako koncepty, přestože nemají své vlastní instance.

Uvažujme například *Parcely*, na nichž roste *Energetická rostlina*. Tyto *Parcely* by se z našeho pohledu mohly lišit zeměpisnou polohou (např. v rámci ČR – severní a jižní), v takovém případě bychom mohli uvažovat o dalších podkonceptech *Severní parcela* a *Jižní parcela*. U naší ontologie však plně postačuje koncept *Parcela* (viz obr. 4), která může mít pouze určité instance jako *Parcela_OH_2009_N* nebo *Parcela_COOH_2010_NO*.



Obrázek 4. U konceptu *Parcela* postačí instance (neuvažujeme další podkoncepty).

11.6 Limitace oblasti

Ontologie nemusí obsahovat všechny dostupné informace o doméně – není potřeba specifikovat (nebo zobecnit) více, než by mělo a mohlo být potřeba pro její možnou aplikaci (obyčejně nejvýše jednu extra úroveň). Pro ontologii huminových látek není například nutné znát materiál nástrojů pro odebrání vzorků nebo věk člověka, který vzorky odebíral. Ontologie nemusí obsahovat rovněž všechny role a omezení hierarchie konceptů. Do naší ontologie rozhodně nebudeme zahrnovat všechny možné role prvků a sloučenin, které mohou mít. Vybereme jen nejvýznačnější role konceptů v ontologii, ty, které pomohou k zodpovězení otázek kompetence a ty, které považuje za nutné doménový expert. V ontologii nebudou také nutně zahrnuty všechny vztahy, které si lze představit mezi elementy definovanými v ontologii, zejména takové, které jsou irelevantní vzhledem ke kontextu námi řešených experimentů. Tato návrhová rozhodnutí je vhodné zahrnout do dokumentace ontologie, což usnadní pochopení uživatelům ontologie.

11.7 Disjunktní koncepty

U většiny systémů pro tvorbu ontologií lze explicitně specifikovat disjunktnost konceptů. Koncepty jsou disjunktní tehdy, když nemohou mít žádné společné instance. Například *Výsledek_měření* a *Výsledek_pokusu* nebudou disjunktní, protože bychom mohli například najít takové výsledky pokusů, které se shodují s výsledky měření.

11.8 Inverzní vlastnosti

Hodnota role může záviset na hodnotě jiné role. Například *Substrát je_na Parcele* a *Parcela obsahuje Substrát*. Tyto dvě role jsou inverzní. Ukládání rolí v obou směrech je redundantní. Víme-li, že *Substrát je_na Parcele*, pak aplikace, která použije naši ontologii, dokáže odvodit hodnotu inverzní relace. Pro naši kontrolu správnosti definovaných vztahů je však někdy výhodné, aby byly doplněny i inverzní role. Tento přístup dovolí uživateli určit konkrétní *Parcelu* v jednom případě a *Substrát* v jiném, definici konceptů lze přirozeně „přečíst oběma směry“.

11.9 Doporučené hodnoty

Mnoho systémů založených na rámcích (např. Protégé-Frames [18]) dovoluje specifikovat doporučené hodnoty rolí. Jestliže je hodnota role stejná pro většinu instancí konceptu, je možné ji definovat jako výchozí. Potom, je-li vytvořena instance obsahující tuto roli, je tato hodnota vyplněna automaticky, je ji ale možné změnit na jakoukoli jinou povolenou hodnotu, nepřináší tedy žádné nové omezení.

11.10 Jmenné konvence

Definice a pozdější dodržování jmenných konvencí v ontologii je důležitým krokem k jejímu snazšímu pochopení a často rovněž zabrání běžným chybám při modelování. Pojmenovávání konceptů a rolí má jistě více alternativ. K výběru jedné z nich většinou není přesný důvod, ale je výhodné určit jmennou konvenci pro koncepty a role, které se budeme při návrhu držet. Určitá pravidla jsou často určena i výběrem systému pro reprezentaci znalostí (např. Protégé [17]). Následující rysy takových systémů ovlivňují výběr jmenné konvence:

1. Má systém stejný jmenný prostor pro koncepty, role a instance? Tj. dovoluje systém mít koncept a roli stejného jména?
2. Je systém "case-sensitive"? Tj. rozliší systém jména konceptů, která se liší jen ve velikosti písmen?
3. Jaké oddělovače systém ve jménech konceptů dovoluje? Mohou jména obsahovat spojovníky, podtržítka, mezery, vykřičníky apod.?

Je dobré rozhodnout tyto další detaily jmenných konvencí a ty se snažit dodržovat v celé ontologii:

1. určení oddělovače a velikosti písmen (*ChemickyProces*, *Chemický_proces*, *chemický-proces*, *Chemický proces*, ...),
2. jednotné či množné číslo u názvů konceptů (*Chemický_proces* nebo *Chemické_procesy*?),
3. předpony (*je_*, *má_*, ...) či přípony (*_na*, *_do*, ...) názvů rolí.

Vhodné je rovněž nedávat jménům konceptů a rolí předpony a přípony *koncept* (popř. *třída*), *role* (popř. *slot*), a nezkracovat názvy konceptů.

12 Závěr

V této práci jsme popsali postup vývoje ontologie týkající se huminových látek. Snahou bylo popsat pravidla pro tvorbu takové ontologie a odůvodnit navrženou strukturu (viz obr. 2, 3 a 4). V jednotlivých krocích byly popsány (obecné) požadavky pro vytvoření hierarchie konceptů, rolí a jejich hodnot. Ačkoli jsme se snažili navrhnout ontologii podle obecných pravidel a podle nejlepšího vědomí, je potřeba znovu upozornit na to, že *neexistuje jediná správná ontologie pro určitou doménu*. Její návrh je tvůrčím procesem a dvě ontologie, které popisují stejnou doménu, ale jsou navržené dvěma různými skupinami lidí, nemusí být stejné. Potenciální aplikace ontologie bezpochyby ovlivní postupy při jejím navrhování. Kvalitu naší ontologie lze zhodnotit použitím v aplikaci, pro níž byla ontologie navržena.

Reference

1. M. Grüninger, M. S. Fox: *Methodology for the Design and Evaluation of Ontologies*. In: Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal (CA), 1995.

2. M. Uschold, M. Grüninger: *Ontologies: Principles, Methods and Applications*. In: Knowledge Engineering Review 11(2), Edinburgh (UK), 1996.
3. M. A. Wilson, M. A. Vassalo, E. M. Purdie, J. H. Reuter: *Compositional and solid-state nuclear magnetic resonance study of humic and fulvic acid fractions of soil organic matter*. In: Analytical Chem. 59, s. 551–558, 1987.
- [I1] N. F. Noy, D. L. McGuinness: *A Guide to Creating Your First Ontology*. Online: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf.
- [I2] DAML – DARPA Agent Markup Language. Online: <http://www.daml.org>.
- [I3] HCLS – Semantic Web Health Care and Life Sciences Interest Group. Online: <http://www.w3.org/2001/sw/hcls>.
- [I4] ChEBI – Ontology for Chemical Entities of Biological Interest. Online: <http://www.ebi.ac.uk/chebi>.
- [I5] Ontolingua – Databáze a ontologie. Online: <http://www.ksl.stanford.edu/software/ontolingua>.
- [I6] OWL – Web Ontology Language / W3C Semantic Web Activity. Online: <http://www.w3.org/2007/OWL>.
- [I7] Protégé – Ontology Editor and Knowledge Acquisition System. Online: <http://protege.stanford.edu>.
- [I8] Protégé-Frames – Frame-based Domain Ontology Editor. Online: <http://protege.stanford.edu/overview/protege-frames.html>.
- [I9] SWOOP – Hypermedia-based OWL Ontology Browser and Editor. Online: <http://code.google.com/p/swoop>.
- [I10] UNSPSC – The United Nations Standard Products and Services Code. Online: <http://www.unspsc.org>.
- [I11] W3C – World Wide Web konsorcium. Online: <http://www.w3.org>.