

DIGITALIZACE PRÁVA – ACADEMICLEX¹

František Cvrček

Program AcademicLex

V Ústavu státu a práva AV ČR (ÚSP) v sekci Centrum pro výzkum právního jazyka (CVPJ) vyvíjíme program AcademicLex (ALX) od roku 2019.² ALX je podporován Strategií 21 Akademie věd ČR. Na programu spolupracují právníci ÚSP, programátoři Fakulty informatiky MU v Brně a programátoři firmy C-soft, učitelé Právnické fakulty ZČU v Plzni a další pracovníci zabezpečující zpracování dat. ALX navazuje na projekt PES (Právnický elektronický slovník)³ taktéž dlouhodobě podporovaný Strategií 21 AV ČR. Do systému jsou začleněny právní předpisy ČR, kompletní od roku 1918, judikatura a důvodové zprávy v celkovém rozsahu 200 000 dokumentů v registru a přes 100 000 v plných textech. Data jsme shromažďovali 30 let a systém průběžně aktualizujeme.

ALX je originální program kompletně vytvořený v ÚSP na základě výzkumů českého právního jazyka. Speciální slovník získaný analýzou právních textů obsahuje 850 000 slov. Slovník umožňuje nejen lematizaci, ale také generuje sémantické skupiny, které řeší přechody mezi slovními druhy, prefixy, sufixy, morfologické

¹ Článek byl vytvořen s podporou Strategie AV21: RVO: 68378122.

² Viz CVRČEK, F. *Manuál AcademicLex*. E-kniha. Praha: Ústav státu a práva AV ČR, 2020, 75 s. Dále CVRČEK, F., HAVLIN, D. *AcademicLex*. Dostupné z: <http://academiclex.ilaw.cas.cz/download/Praha>.

³ Viz CVRČEK, F., PALA, K., RYCHLÝ, P. *Legal Electronic Dictionary fr Czech. Proceedings of the Eight International Conference on Language Resources Association (ELRA)*. Istanbul, 2012. Dále CVRČEK, F., RYCHLÝ, P., KOCINCOVÁ, L. *Právnický elektronický slovník*. Verze 2016. Dostupné z: <https://deb.fi.muni.cz/pesnew>.

změny v gramatice, starší česká slova apod.

Každý dokument je zpracován podle pevné normy, což umožňuje právní dekompozici a možnosti vyhledávání podle volby struktury, např. normativní text, poznámky, rubriky, formální části (hlavy, oddíly, písmena apod.), právní věty judikátů atd. Vedle dekompozice právní je k dispozici i dekompozice gramatická.

Na základě analýz jsme zjistili, že k interpretaci právních předpisů každý právník pracuje s doktrinárními termíny, obvykle v poměru 1 : 5, tj. na jeden termín právního předpisu se u klasických odvětví používá až 5 termínů doktrinárních. To nás vedlo k postupům dekomponujícím právní učebnice a knihy (tzv. **doktrinární texty**) a tvorbě speciálních funkcí pro vyhledávání a automatické rozpoznávání doktrinárních struktur.

ALX zachovává všechny funkce předcházejících programů (CS LEGSYS, LEGSYS, LexGalaxy) a rozvíjí nové postupy pro analýzu, vyhledávání a zpracování doktrinárních textů.

Zpracování doktrinárních textů

Vzhledem k respektování autorského práva se zabýváme pouze texty starými (autorsky volnými) nebo texty současnými, ke kterým máme souhlas autorů.

Staré texty jsou slovníkově zpracovány programem OCR a zkontrolovány ověřovacími postupy. Poté se provede kontrola ve Wordu a následuje kontrola proti našim speciálním právním slovníkům. Přes všechny tyto kontroly nakonec stejně musí poslední kontrolu provést specialista z daného oboru a udělat korekturu. Prostě standardní programy OCR jsou sice dobrou podporou, ale jejich chybovost je natolik velká, že bez klasické korektury se dobře zpracovaný text neobejde.

Co se týká struktur, jejich rozmanitost je u doktrinálních textů natolik variabilní, že ani nemá smysl tiskové značky (Hrt, Indent, TAB atd.) automaticky rozpoznávat. Je lepší použít normu, podle které se tiskové značky do textu vloží, při respektování autorova záměru.

Klasické texty, např. Heyrovský⁴, Sommer⁵, Arndts⁶, Sedláček⁷, Kallab⁸ atd., jsou obvykle velmi dobře strukturovány, tj. jak co do členění, tak struktury rubrik apod. Platí zásada, že zpracovatel může do textu vložit podrobnější členění či zvýraznění, ale jeho vklad je jednoznačně odlišitelný od úvodníku autora. Text tak může být bohatší, ale nikdy nesmí jakkoli deformovat původní text. Chce-li uživatel číst pouze čisté autorské znění bez vložených informací, má tuto možnost. To ovšem znamená, že musíme strukturu textu variabilně přizpůsobovat požadavkům uživatelů.

Předpokládáme, že doktrinální texty se skládají z termínů (atomárních a složených), frází a vět či pasáží významných, např. protože jde o definici apod. První problém nastává, když požadujeme, aby specialista (obvykle učitel daného oboru) **označil atomární termíny daného textu**. Jediné omezení spočívá v počtu termínů. Např. podle konvence jsme ohraničili počet atomárních termínů, které by měl student zvládnout za jeden semestr, na 500. Pokud jde o přednášky, kterých je obvykle 10, max. 12, zhruba připadá na jednu přednášku 50 atomárních termínů. Je to samozřejmě požadavek, který není třeba dodržet přesně, ale student by neměl být vystaven představě, že za semestr má zvládnout základních termínů např. 1000 a více. Vlastně jde o dohodu učitelů dané katedry, co považují za minimum pro absolvování zkoušky. Taková dohoda není jednoduchá, ale pokud

⁴ HEYROVSKÝ, L. *Dějiny a systém soukromého práva římského*. Díl I a II. Praha, 1910.

⁵ SOMMER, O. *Soukromé právo římské*. Praha, 1946. Dále SOMMER, O. *Prameny soukromého práva římského*. Praha, 1932.

⁶ ARNDTS, C. L. *Učební kniha pandekt*. Díl I, II a III. Praha, 1886.

⁷ SEDLÁČEK, J. *Občanské právo československé*. Brno, 1931.

⁸ KALLAB, J. *Úvod ve studium metod právnických*. Díl I a II. Brno, 1920.

se škrtá z nabídnutého seznamu, lze tento požadavek splnit. Za současné situace se předpokládá, že student má zvládnout 3krát více termínů, ačkoliv je jasné, že to prostě není možné.

Je pravda, že klasické učebnice obsahovaly několik tisíc atomárních termínů, ale strukturálně bylo zřejmé, které termíny jsou podstatné (viz např. Sommer). Učebnice byla sice velmi bohatá, obvykle provázaná s jinými texty, ale také snadno redukovatelná. Bohužel současné učebnice splňují tyto vlastnosti pouze výjimečně.

Druhý problém souvisí s **linearitou textu**. Tento problém se objevil již v projektu PES, který byl založen na hierarchicky uspořádané struktuře rubrik. Autor doktrinálního textu obvykle předpokládá, že čtenář bude číst text průběžně od začátku do konce, tedy postupuje jakoby po úsečce. S dobře digitalizovaným textem se ovšem pracuje odněkud odlišně. Čtenář (uživatel) vyhledává dílčí část textu a posléze se jí snaží obklopit souvisejícími informacemi. Pohybuje se spíše na ploše nebo ještě lépe v prostoru. Pokud autor od počátku tvorby textu nebere v úvahu, že text bude digitalizován (tedy promítnut do prostoru), a setrvá v klasické linearitě, vznikne mnoho nejasností s přechodem od úsečky do prostoru. Starší autoři si patrně uvědomovali, že tento problém může nastat, a čelili mu jednak bohatou strukturou rubrik, včetně formálního členění, a jednak několika typy zvýrazňování částí textu.⁹ Pro někoho možná překvapivě jsou starší klasické učebnice z hlediska inforatického zpracování mnohem kvalitněji než ty současné. Důvody tohoto jevu nejsou sice zcela jasné, ale naskýtá se hypotéza, že svému oboru skutečně dobře rozuměli. Pokud budeme chtít využít současných možností digitalizace, patrně budeme muset vytvořit nové doktrinální texty, které budou od počátku s následnou digitalizací počítat.

⁹ Za klasicky dobře zpracovaný text považujeme Bibli (jeruzalémskou nebo šestidílnou kralickou).

Ne nepodstatný vliv na tvorbu digitálních doktrinálních textů mají **hypertextové odkazy**. Jestliže ALX obsahuje např. základní díla právní filosofie (Platon¹⁰, Aristoteles¹¹, Dante¹², Machiavelli¹³, Hobbes¹⁴, Montesquieu¹⁵, Rousseau¹⁶ atd.), přednášky se mohou opírat o přímé odkazy na příslušné místo včetně možnosti konfrontace s originálním zněním. Tím máme k dispozici širší kontext a snižuje se tak možnost nepřesné interpretace.

Každý starší text **obohacujeme o strukturální členění** jak formální, tak obsahové, které do textu vnoří zpracovatel. Formálním členěním se miní paragrafy, pokud text na ně nerozčlenil sám autor, a případně nižší jednotky (odstavce apod.). Obsahovým členěním miníme rozlišení atomárních termínů, nejdůležitějších frází, doplňujících frází, kritických frází atd. Digitalizovaný text lze pak zobrazovat podle vybrané úrovně, např. uživatel chce vidět pouze nejdůležitější fráze nebo jen přehled atomárních termínů, případně kritické fráze ve struktuře rubrik apod. Vyhledávání lze pak zaměřit na kteroukoliv ze zvolených struktur, např. hledám termín „spravedlnost“ pouze v nejdůležitějších frázích.

Stručně shrnuto, pokud budeme brát digitalizaci práva vážně, nelze se obejít bez **podrobné analýzy dat a jejich zpracování s tím, že programy mají být přizpůsobeny typu daných textů**. Představa, že vytvoříme obecný SW pro všechny typy textů, je iluzorní. Jestliže chceme vytvořit učebnice a přednášky pro studenty právnických fakult, budeme je muset znovu napsat s tím, že si budeme od počátku vědomi, že naše texty budou plně digitalizovány a bude se s nimi takto pracovat. Domnívá-li

¹⁰ PLATON. *Ústava*. Praha, 1921. Dále PLATON. *Zákony*. Praha, 1925.

¹¹ ARISTOTELES. *Politika*. Praha, 1895.

¹² DANTE, A. *O jediné vládě (Monarchie)*. Praha, 1942.

¹³ MACHIAVELLI, N. *Vladař*. 1940.

¹⁴ HOBBS, T. *Leviatan neboli o podstatě, zřízení a moci státu církevního a občanského*. Praha, 1941.

¹⁵ MONTESQUIEU, CH. *O duchu zákonů*. Praha, 1947.

¹⁶ ROUSSEAU, J. J. *O společenské smlouvě neboli o zásadách státního práva*. Praha, 1949. Dále ROUSSEAU, J. J. *O původu nerovnosti mezi lidmi*. Praha, 1949.

se někdo, že digitalizace spočívá v tom, že napíše text ve Wordu a vystaví jej na WEB, nemá tušení, o čem mluví.

Koncepce ALX

První vlna pandemie nás zastihla v druhém roce vývoje programu ALX, což významně ovlivnilo další postup. Výchozí koncepce byla založena na předpokladu, že **student by měl mít k dispozici všechny informace nutné ke zkoušce**, tj. přednášky, učebnice, klasická díla z daného oboru, právní předpisy, judikaturu, důvodové zprávy atd., v jednom programu, tj. v ALX.

Uvedené texty by měly být vzájemně provázány **hypertextovými odkazy**, včetně odkazů na originály klasických děl, např. Platonova Ústava odkazuje na řecký originál, který je v projektu Perseus velmi dobře zpracován včetně anglického překladu. Student tak může konfrontovat všechny jazykové mutace. Vzhledem k tomu, že ALX obsahuje i rakouské přepisy platné na území Koruny české před rokem 1918, jsou učebnice vydávané na přelomu století a za první republiky provázány s tehdy platným právem (Heyrovský, Arndts, Weyr¹⁷, Sedláček, Kallab atd.).

Z hlediska výběru jsme se soustředili na **obory**: teorie práva, pandektní právo, římské právo, právní filosofie, legislativa a právní informatika.

Všechny doktrinální texty jsou zpracovány podle **jednotné normy**, obohaceny zpracovateli o označení atomárních termínů, složených termínů, frází a rubrikové struktury.

Vyhledávání lze omezit na libovolně zvolenou **strukturu**. Každý text lze „řezat“ podle zvolené struktury.

¹⁷WEYR, F. *Teorie práva*. Brno, 1936.

U **nových textů** se autoři snaží respektovat požadavky kladené na dobře digitalizovaný text, což mnohdy není zdaleka jednoduché. Je totiž třeba mít zcela jasnou představu o struktuře daného oboru, termínech, odstavních frázích a definicích s přihlédnutím k pedagogickému aspektu u učebnic a přednášek.

Systém ALX mají **studenti k dispozici zdarma**. Stáhnou si jej z úložiště v ÚSP a mohou s ním libovolně pracovat **bez závislosti na Web**. Samozřejmě připojení na Web obohatí ALX o informace z řady dalších bází, jako je Perseus, Gutenberg, báze Senátu, Poslanecké sněmovny, Ministerstva vnitra, Nejvyššího soudu, Ústavního soudu atd.

Přednášky jsou doplněny **videi** zpracovanými v programu FlashBack a jsou součástí ALX.

Principiálně by studenti měli být bez problémů schopni složit zkoušky z výše uvedených oborů, pokud se umí samostatně učit. Učitelé byli ve spojení se studenty a poskytovali konzultace, pokud o ně studenti měli zájem. Vedle systému ALX řada učitelů využívala obecné SW pro výuku online, takže máme docela jasnou představu o problémech spojených s výukou v době pandemie, kdy bylo de facto zrušeno kontaktní studium v řadě oborů.

Zkušenosti s digitalizací práva na PF v době pandemie

Základní zásada zní: **Digitalizace má smysl pouze tehdy, pokud povede ke zvýšení kvality výuky.**¹⁸

Zkušenosti s výukou v průběhu zrušení kontaktní výuky ukázaly, že **využití obecných SW** umožňujících nahrávání přednášek

¹⁸ CVRČEK, F. Několik poznámek k digitalizaci výuky práva. In: *Metamorfózy práva ve střední Evropě 2020: Právo a krize*. Plzeň, 2020.

vede k degradaci výuky. V dané situaci je to sice dobrá náhražka, ovšem kontaktní výuku nenahradí. Ještě markantnější je tento trend v případě seminářů.

Efektivnější se jeví specializované **programy pro jednotlivé obory**. Příkladem takových SW je ALX nebo PES. Výše jsme uvedli obory, které pokrývá program ALX. U každého oboru bylo třeba položit důraz na jiný typ prezentace. U právní informatiky je jádro postaveno na videích, která manifestují jednotlivé postupy vyhledávání a řešení konkrétních úloh s komentářem. Písemné přednášky stručně shrnují teoretická východiska a základní informace z učebnice. Naproti tomu u právní filosofie představují hlavní osu písemné přednášky, které podporují odkazy na české překlady klasických děl a je možná konfrontace se zněním v původním jazyce. U legislativy stojí obor na učebnici, která odkazuje na videa, šablony a podpůrné texty typu Legislativních pravidel vlády, RIA apod.

U speciálních SW nejde pouze o volbu formálních prezentací, ale především o přizpůsobení obsahových struktur, atomárních termínů a frází danému oboru. Rozhodující je předmět a nelze podléhat diktátu metody, nebo dokonce techniky. Tvorba takovýchto specializovaných SW je dlouhodobá, náročná a drahá. Nelze předpokládat, že PF by byly schopny samy takové SW vyprodukovat. ALX a PES byly vyvinuty v AV ČR s podporou Fakulty informatiky MU. PF se podílely na zpracování dat. Bez spojení více akademických pracovišť a dlouhodobé podpory (v našem případě AV ČR) nejsou takové projekty reálné. Představa, že rozsáhlé báze a speciální SW lze vytvořit na základě podpory 3- až 5letým grantem, jsou z oblasti naivních iluzí.

Výhodou specializovaných SW s rozsáhlými bázemi je prezentace oboru jako složité struktury, kdy **informace nutné ke zkoušce představují pouze jednu vrstvu**. Strukturální uvažování je v právu prioritní, neboť právo je především relační složité

struktura. ALX ukazuje tuto složitost, ale umožňuje určit hranice minimálních znalostí nutných ke zkoušce.

Nemá smysl zastírat, že i přes veškerou snahu **specializované SW naráží na řadu problémů**. První problém spočívá v tom, že studenti nejsou zvyklí samostatně studovat. Dále mají tendenci zjednodušovat. Značné nedostatky spočívají v jazykové nevybavenosti, většina vládne pouze angličtinou. Neporozumění složitějším textům je více méně pravidlem. U starších textů se setkáváme s neporozuměním starším verzím češtiny. Způsoby vyhledávání v informačních systémech jsou zatíženy primitivními technologiemi, které jsou propagovány tzv. trhem. Je třeba si uvědomit, že komerční právní informační systémy od 90. let neučinily žádný významnější pokrok. Zjednodušeně shrnuto, dobrý SW zefektivní studium dobrým studentům, ale průměrným a špatným příliš nepomůže. Otázkou ovšem je, zda skutečně potřebujeme tolik průměrných a podprůměrných právníků. Jestliže jsme schopni studentům poskytnout kvalitní informace, domnívám se, že bychom měli také **nastolit přísnější kritéria**.

Výše jsme se zmínili o nutnosti přepsat učebnice ve světle možné digitalizace. Něco podobného se týká i kontaktního studia, především přednášek. Naprostá většina učitelů se shoduje, že kontaktní studium je nenahraditelné, a to se týká jak příznivců digitalizace, tak jejich odpůrců. V čem je ovšem ona nezastupitelnost, už jasné příliš není. U některých učitelů převládá strach z toho, že když svoji přednášku natočí na video a zveřejní, nebudou už mít co přednášet. Obvykle se tento postup schovává za autorská práva. Tyto názory ponechme stranou. Faktem ovšem je, že možnosti prezentace v rámci specializovaných SW předpokládají dost radikální změny nejen přednášek, ale i **změny koncepce celého studia**. Nemá smysl si nalhávat, že specializované SW vše vyřeší. Učitelé se také učí s nimi pracovat, SW se zdokonalují pomalu a panuje zde mnoho nejistot. Jisté je, že po dlouhou dobu digitalizace výuky práva poběží

paralelně s klasickým kontaktním studiem. V každém případě je digitalizace neodvratný proces. Problémem ovšem je, abychom postupovali uvážlivě a postupně podporovali formy, které kvalitu studia zlepšují. Zatím se zdá, že o digitalizaci práva se na PF sice hodně mluví, ale reálnou podporu nikde nevidím.¹⁹

Při zpracovávání klasických právnických učebnic přelomu 19. a 20. století mám někdy pocit, že tito autoři věděli o digitalizaci práva více než ti současní. Alespoň z organizace informací to tak vypadá.

Osobní poznámka

Právní informatikou se zabývám v ÚSP 37 let. Když jsem v roce 2010 dopisoval monografii „Právní informatika“,²⁰ která shrnovala výsledky výzkumu od roku 1985, požádal jsem o připomínky dr. Z. Masopusta. Znal jsem jej od roku 1981, kdy jsem do ÚSP přišel studovat aspiranturu u prof. V. Knappa. Dr. Masopust pak byl mým vedoucím v oddělení teorie práva. Dále byl vedoucím Katedry teorie práva ZČU, výkonným redaktorem *Právnicka*, překladatelem z angličtiny, němčiny a ruštiny, erudovaným autorem mnoha článků, monografií a expertiz. Pro mě ale především byl mým přítelem, kterého jsem si vážil. Ačkoliv se nevěnoval právní informatice, obrátil jsem se právě na něj, protože jsem chtěl, aby mně někdo řekl, zda je můj text česky, zda je srozumitelný a smysluplný. Všechny poznámky, doplňky a změny, které dr. Masopust navrhl, jsem bez výjimky akceptoval, protože on to prostě uměl. Společně jsme založili cyklus konferencí *Metamorfózy práva*, které po léta moderoval. Vzpomínám na něj jako na humanistu, demokrata a nesmírně vzdělaného člověka. Jsou lidé, kteří nám chybí.

¹⁹ To neplatí o AV ČR, která digitalizaci vždy reálně podporovala.

²⁰ CVRČEK, F. *Právní informatika*. Praha, 2010.