

Poznávání rostlin po našem: co prozradila mobilní aplikace

Analytická data získaná během provozu aplikace FlowerChecker umožňují odpovídat na otázky ohledně zastoupení typů dotazovaných rostlin (bližší na str. 121 této Živy) i nejčastějších taxonů. První místo zaujímá rod pryšec (*Euphorbia*), jeden z nejbohatších rostlinných rodů. Následuje překvapivě rod kolopekka (*Kalanchoe*), který zahrnuje několik pěstovaných sukulentů z tropů a subtropů, a rod dračinec (*Dracaena*). Nasbíraná data v poslední době využíváme i jiným způsobem. Trénujeme jimi neuronovou síť, která má za úkol rostliny rozpoznávat automaticky. Už nyní je schopna odlišit více než 8 tisíc taxonomických tříd a v budoucnu jistě najde uplatnění v různých odvětvích, především v zemědělství a ochraně přírody.

Na konci léta 2013 se sešlo několik okolností, které odstartovaly vývoj aplikace FlowerChecker na poznávání rostlin. V Brně vrcholila jedna z prvních startupových vln, v odborných člancích zabývajících se strojovým učením neuronových sítí se začal ukazovat potenciál, který skrývají metody hlubokého učení (též např. Živa 2014, 5: CXIII–CXIV). Sestěhovali se také studenti – dva z budoucích zakladatelů aplikace, botanik Ondřej Vild a programátor Ondřej Veselý. Společné rozhovory se stočily i k tématu poznávání rostlin. Uvědomili jsme si, že pro laika je zjištění druhu neznámé rostliny velmi obtížný úkol, který nelze jednoduše vyřešit „googlováním“. Vznikl nápad využít informační technologie k tomu, aby se lidé při určování neznámých rostlin mohli propojit s botaniky.

Pro automatické poznávání rostlin z fotografií tehdy neexistovala dostupná technologie. Přesto bylo poměrně zřetelné, že tahouni pokroku v oboru, velké soukromé firmy, začnou pro strojové učení nabízet své technologické platformy. Uvědomili jsme si, že data získaná expertní identifikací budeme moci použít pro trénování neuronové sítě, jakmile bude technologie dostupná.

První funkční prototyp naprogramoval Ondřej Veselý doslova přes víkend. Měl odpovědět na otázku, kdo je naším potenciálním uživatelem. Vyvinuli jsme jednoduchý software a připojili ho na aplikační rozhraní sociální sítě Twitter, aby prostřednictvím naivního bayesovského filtru monitoroval různé varianty otázky „Nevíte někdo, prosím, jaký je název této kytky?“ (a to v různých jazycích). Přiložené snímky se

pak objevovaly v prohlížeči tehdejšímu doktorandovi Vildovi, který na otázku ve většině případů dokázal uspokojivě odpovědět. Uživatelé byli nadšeni, že jim někdo pomohl, a často byli ochotni i vyplnit dotazník. Za několik týdnů se tak zakladatelé dozvěděli, že v naprosté většině případů lze rostliny z fotografie určit dostatečně přesně, poptávajícími bývají nejčastěji zahrádkáři z USA a přibližně desetina z nich je ochotna za službu zaplatit. Dospěli jsme k závěru, že můžeme začít budovat databázi rostlin a být za tuto práci dokonce finančně kompenzováni. Proto do týmu vzápětí přibyl Jiří Řihák, který se pustil do vývoje mobilní aplikace, přes níž už za několik měsíců uživatelé poslali stovky dotazů týdně. Začalo být zřejmé, že je nutné posílit i zastoupení botaniků.

Ze začátku tvořili skupinu určovatelů pouze další studenti botaniky na Přírodovědecké fakultě Masarykovy univerzity, později se ale podařilo oslovit nadšence z různých koutů světa. Do týmu přibyla knihovnice z Baltimoru, zahradnice z květinářství v Severní Karolíně, bývalá právnička z australského východního pobřeží nebo indický emeritní profesor taxonomie. Za současné složení přibližně 20členného týmu by se tak nemusela stydět žádná společnost, které záleží na různorodosti.

Kdo aplikaci využívá?

Aplikace je nyní dostupná jen v Evropě, Severní Americe, Austrálii a Indii. Za posledních pět let přišlo téměř 250 tisíc žádostí o identifikaci rostlin. Není žádným překvapením, že největší podíl požadavků pochází ze Spojených států amerických (viz obr. 3 na str. 121). Míra dominance

- 1 Pryšec zářivý (*Euphorbia milii*) známý též jako Kristova koruna je dřevnatý sukulent s hustými trny na lodyze a nápadnými, nejčastěji červeně zbarvenými listeny, podpírajícími miniaturní redukované květenství.
- 2 Jeden ze zachycených tweetů



však předčila očekávání. Vysvětlujeme si ji různými mentalitami národů. Zatímco lidé ze starého kontinentu si stále ještě raději otevřou knížku nebo se zeptají souseda, Američané jsou zvyklí hledat pomoc na internetu i s řešením otázek typu: „Co mi to nového vyrostlo na zahradě? Mám tu rostlinu vytrhnout, nebo ji zalévat?“ Určitou roli zároveň může hrát i vyšší kupní síla americké měny, protože cena za využití databáze se mezi regiony téměř neliší.

Nejčastěji určované rostliny

Zpočátku FlowerChecker sliboval jen určování cévnatých rostlin. Později jsme se sice odvážili i na další skupiny organismů, přesto nyní tvoří mechorosty, houby a lišejníky jen asi 2 % určených objektů a v tomto článku se jim dále nebudeme věnovat. Současná databáze obsahuje téměř 9 tisíc druhů cévnatých rostlin. Jde asi o pouhých 2 % dnes známých druhů, přesto poskytuje zajímavý pohled, jak jsou různé rostliny časté a které přitahují nejvíce pozornosti.

Když uživatelé nahrávají fotografie, mohou je zařadit do některé z kategorií (pokojová, zahradní či planá rostlina), aby pomohli s identifikací. Lze tak porovnat, jaké typy rostlin většinou hledají. Je zajímavé, že vnímání jednotlivých druhů se mění v průběhu roku. Třeba gerbery a hyacinty uživatelé řadí buď mezi pokojové, nebo zahradní – zatímco během února jsou hyacinty považovány výhradně za zahradní, od května už bývají označovány převážně za pokojové. Z celkového zastoupení v 60 % dotazů jde o rostliny pěstované, přičemž venkovní výrazně převládají nad pokojovými. Naproti tomu plané druhy (včetně pleveľných v zahradách) bývají předmětem dotazu jen přibližně ve třetině případů.

Které rody se řadí mezi nejčastěji dotazované? Pro botaniky nebude překvapením, že zvítězil rod pryšec. Je totiž jedním z největších a nekomplexnějších rodů, od kterého jsme dosud nashromáždili úctyhodných 100 druhů. Mezi nejběžnější patří řada pěstovaných – např. pryšec nádherný, známý také jako vánoční hvězda (*E. pulcherrima*), nebo p. zářivý zvaný Kristova koruna (*E. milii*, obr. 1). Velmi často se objevuje pryšec skočcový (*E. lathyris*, obr. 1 na str. 121), dvouletá bylina původem z jižní Evropy, která se pro zajímavé listy pěstuje v mírném pásmu celého světa, ale mnohde uniká do volné přírody a stává se invazní.

Druhým nejčastějším je rod kolopejka z čeledi tlusticovitých (*Crassulaceae*), s centrem rozšíření na Madagaskaru. Pěstuje se řada druhů, avšak na rozdíl od pryšců patří veškeré zásluhy jedinému z nich – kolopejce vděčné (*K. blossfeldiana*). Nenáročný sukulent je s 1 600 dotazy suverénně naším nejpoptávanějším druhem. Najdeme ji sice v každém květinářství, přesto ale zůstává mezi lidmi méně známá, byla totiž kultivována relativně pozdě. Nese jméno německého zahradníka Roberta Blossfelda, který kolopejku mezi světovými válkami přivezl do Evropy. Ale až v 80. letech 20. století si druh získal popularitu díky vyšlechtění mnoha kultivarů.

Na pomyslném třetím místě se umístil rod dračinec. Tyto dřeviny s tuhými listy mají těžiště výskytu v Africe. V pokojových podmínkách kvetou jen zřídka, pěstují se spíše pro pestré listy. Mezi dotazy v aplika-

Plant.id

ci FlowerChecker převažují dva druhy – dračinec vonný (*D. fragrans*) a *D. marginata* (obr. 3). Jde o obecně nenáročnou dřevinu připomínající palmy, ale kvůli obsahu jedovatých saponinů se nedoporučuje jejich pěstování v domácnostech s malými dětmi a domácími mazlíčky.

Je třeba podotknout, že zmíněné statistiky jsou do jisté míry ovlivněny tím, co je tím botanikům FlowerCheckeru schopen identifikovat. Špatná kvalita fotografií nebo stav rostlin může určení i výrazně ztížit. U více než 85 % rostlin byl tím schopen poznat alespoň rod. Z toho se dalo přiřazení zpřesnit do druhového názvu téměř ve čtyřech z pěti případů.

Přechod na automatické určování

V létě 2016 vydala společnost Google softwarovou knihovnu TensorFlow. Ta zavádí většinu metod do té doby popsanych pouze v odborných člancích. Pro účely určování rostlin je klíčové především využití tzv. konvoluční neuronové sítě, jejíž architektura zaměřuje některé své vrstvy na identifikaci specifických vzorů ve zpracovávaném obraze. Na základě kombinací těchto vzorů pak další vrstvy řeší klasifikační úlohu – rozhodují, který druh rostliny je s největší pravděpodobností na vstupní fotografii. Obraz však není jediná informace, na jejímž základě neuronová síť rostlinu určuje. Do vstupní vrstvy lze přidat údaj o poloze, nadmořské výšce a času určení nebo také další informace z tzv. Exif metadata fotografie, která mohou obsahovat např.

3 Druh dračince *Dracaena marginata* pochází z ostrovů v Indickém oceánu. Má charakteristické úzké listy s načervenalým okrajem. Snímky z archivu FlowerCheckeru



na jakou vzdálenost bylo zaostřeno nebo zda byl použit blesk.

První reálné použití prototypu postaveného nad zmíněnou knihovnou TensorFlow se nabízel přímo ve webovém rozhraní, které slouží botanikům pro zadávání identifikovaných druhů. Místo textového pole pro zapsání názvu rostliny může botanik jedním kliknutím vybrat z množiny druhů, které mu předpřipravila neuronová síť. Systém tak často urychluje práci. Tím lidských expertů má samozřejmě stále výrazně vyšší spolehlivost i širší škálu druhů, které dokáže determinovat – lidský faktor je tak v aplikaci FlowerChecker nezbytný. Přesto však automat dokáže ve většině případů napovědět, inspirovat nebo ušetřit cenné vteřiny psaní na klávesnici.

Po necelém roce vylepšování modelu dosáhl systém natrénovaný na určování asi 8 tisíc taxonomických tříd úspěšnosti 60 %. Modelu pomáhá, pokud uživatel zašle více fotografií. Systém tedy dokáže dát smysluplnou a užitečnou odpověď ve výrazně většině případů, proto jsme nabídli tuto službu veřejnosti zdarma pod názvem Plant.id na stejnojmenné doméně. V dubnu 2019 získala tato služba ocenění za nápad roku v prvním ročníku soutěže AI Awards v oboru umělé inteligence (AI – Artificial Intelligence). Služba určování rostlin poskytnutá pomocí strojového učení je od expertního dosti odlišná. Je méně přesná, neřekne, že by potřebovala vědět, jak listy voní, a neporadí, že by bylo lepší rostlinu umístit na světlejší místo. Komerčně se orientuje na firmy a službu jim prodává formou API – aplikačního rozhraní umožňujícího identifikaci snadno začlenit do služeb a produktů jiných firem. Ať už by šlo třeba o autonomní robotickou sekačku, která se vyhne pěstovaným rostlinám, nebo o rozšíření do internetového prohlížeče schopné pojmenovat každou zobrazenou rostlinu.

Budoucnost a plány

Do adopce metod strojového učení v elektronickém obchodování, bankovníctví nebo marketingu se v poslední době investovaly obrovské částky a vývoj probíhá velice rychle. V přírodních vědách ale ještě existují nevyužitá příležitosti. Dosud jsme jednali např. s odborníky v muzeích ohledně vývoje systému pro klasifikaci geologických vzorků. Ve spolupráci s Akademií věd ČR plánujeme projekt na určování rostlin, který by kombinoval informaci o morfologických znacích poskytnutou uživatelem a naší neuronovou síť. Kromě toho budeme podávat projekt zaměřený na určování pylových zrn z mikrofotografií. Experimentujeme také s poloautomatickým určováním rostlin ze snímků pořízených nízkou letícími drony. Nejspíše v příštím roce se můžete těšit na mobilní aplikaci Tree Check (viz www.lifetreecheck.eu) – půjde o terénní hru zaměřenou na druhy stromů ve městech a jejich vliv na klima a vznik tepelných ostrovů.

Pozn. redakce: Jeden ze zakladatelů aplikace Mgr. Ondřej Vild, Ph.D., z Botanického ústavu AV ČR obdržel v říjnu 2018 Cenu Akademie věd ČR pro mladé vědecké pracovníky za vynikající výsledky výzkumu, experimentálního vývoje a inovací, a to za práci s názvem Obnova Rostlinné biodiverzity tradičně obhospodařovaných lesů.

Poznávání rostlin po našem: co prozradila mobilní aplikace

Už šestým rokem lze využívat aplikaci pro mobilní telefony FlowerChecker, pomocí níž si můžete nechat určit rostliny od botanických expertů prostřednictvím fotografií. Analýzou čtvrt milionu přijatých požadavků jsme se pokusili odpovědět především na otázku, na jaké rostliny se lidé ptají nejčastěji a proč. Budoucnost poznávání rostlin vidíme v umělé inteligenci a neuronových sítích. Na základě naší databáze fotografií jsme využili principy hlubokého učení (deep learning) k vytvoření programu schopného rostliny identifikovat automaticky.



1

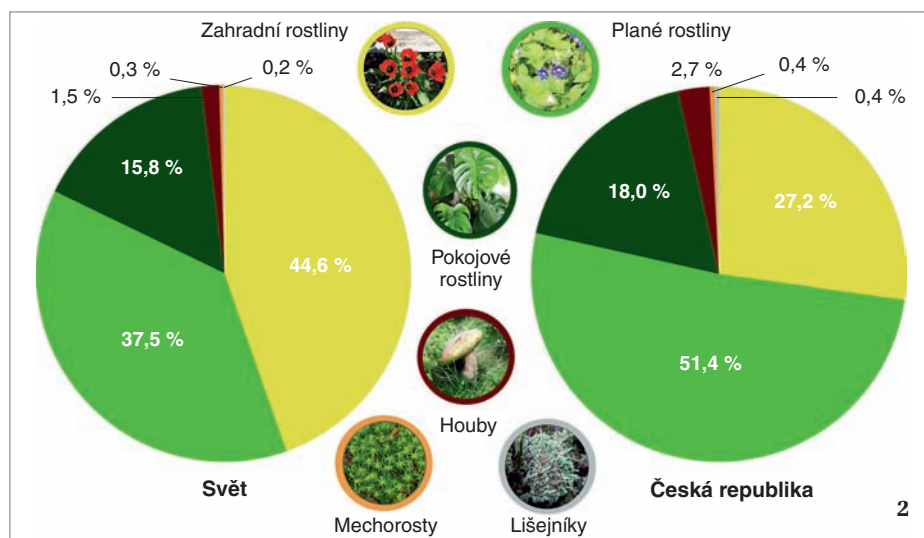
1 Charakteristicky křížmostojné listy pryšce skočcového (*Euphorbia lathyris*), který můžeme často najít v okolí zahrad. Pryšce se v aplikaci FlowerChecker řadí k nejčastěji dotazovaným rodům.

2 Srovnání četnosti požadavků na určité různé typy rostlin, hub a lišejníků napovídá, že se Češi ve srovnání s jinými zeměmi více zajímají o plané rostliny.

3 Podíl počtu požadavků podle států.

Služba určování rostlin je nejoblíbenější v USA. Pochází odtud více než polovina žádostí. Na první pohled překvapí, že počet požadavků z České republiky je vyšší než z Kanady či Německa. Na rozdíl od zahraničních médií se totiž v těch tuzemských informace o aplikaci FlowerChecker již několikrát objevila.

4 Barvy oblastí na mapě ČR se vztahují k počtu identifikovaných rostlin – vzrůstající na škále od bílé po tmavě zelenou. Červeně jsou znázorněny jednotlivé nálezy zákonem chráněných a ohrožených druhů. Z mapy je patrné využití aplikace převážně ve městech a okolí. Chráněné a ohrožené druhy uživatelé nejčastěji posílají z turisticky frekventovaných lokalit. Foto a všechny orig. autoři článku



Srovnání četnosti požadavků z hlediska různých typů rostlin, hub a lišejníků ukazuje, že lidé posílají k určení nejčastěji zahradní rostliny. Podíly četnosti jednotlivých typů rostlin obecně silně závisí na sezoně, nadmořské výšce, zeměpisné šířce a dalších faktorech. Zajímavé je, že třebaže plané rostliny tvoří celkově jen kolem jedné třetiny požadavků, Češi se o ně zajímají výrazně více; zde se jich týká přes polovinu požadavků. Může to souviset s lepší přístupností naší krajiny a relativním bezpeč-

ším pohybem v ní. A to jednak s přístupností v legislativní rovině – vlastník českého lesa nemůže na rozdíl od Spojených států amerických či některých evropských zemí vstup na pozemek zakázat, ani dočasně. Navíc české lesy jsou méně „divoké“ a poměrně přístupné. Ve srovnání s USA nebo Austrálií je u nás daleko nižší pravděpodobnost napadení predátorem a největší riziko zde tak nejspíš představuje klíště obecné. Větší podíl rostlin ze zahrad ve světě je možné vysvětlit ještě jinak. Typickým zahrádká-

řem je totiž starší člověk ze střední třídy a v USA má tato skupina populace obecně větší důvěru k technologiím než u nás. Nízké zastoupení lišejníků, hub a mechorostů bude přirozeně souviset s deklarovaným zaměřením aplikace na cévnaté rostliny. Více o aplikaci FlowerChecker se dozvíte na str. LXXVII–LXXVIII kuléru této Živy.

