PŘIHLÁŠKA VYNÁLEZU

Zveřejněná podle §31 zákona č. 527/1990 Sb.

(21) Číslo dokumentu:

2023-154

(13) Druh dokumentu: **A3**

(51) Int. Cl.:

G10L 13/04 (2013.01)
G10L 13/08 (2013.01)

(19) ČESKÁ REPUBLIKA

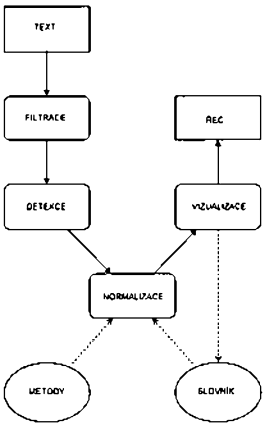
ÚŘAD PRŮMYSLOVÉHO VLASTNICTVÍ

(22) Přihlášeno: **09.11.2021**
(32) Datum podání prioritní přihlášky: **09.11.2021**
(32) Číslo prioritní přihlášky: **PCT/CZ2021/000054**
(32) Země priority: **IB**
(40) Datum zveřejnění přihlášky vynálezu: **13.09.2023**
(Věstník č. 37/2023)
(86) PCT číslo: **PCT/CZ2021/000054**

(71) Přihlašovatel:
Západočeská univerzita v Plzni, Fakulta designu a umění Ladislava Sutnara, Plzeň, Jižní Předměstí, CZ

(72) Původce:
doc. Ing. Jindřich Matoušek, Ph.D., Chlumčany, CZ
Ing. Martin Grůber, Ph.D., Plzeň, Východní Předměstí, CZ
Ing. Markéta Řezáčková, Plzeň, Skvrňany, CZ
Ing. Jakub Vít, Plzeň, Lobzy, CZ

(74) Zástupce:
POLÁČEK a kol., patentová a známková kancelář, Ing. Jiří Poláček, Dominikánská 287/6, 301 00 Plzeň, Vnitřní Město



(54) Název přihlášky vynálezu:

Způsob převodu rozhodnutí orgánu veřejné moci z ortografické do fonetické podoby

(57) Anotace:
Způsob převodu textu rozhodnutí orgánů veřejné moci z ortografické do fonetické podoby pro potřeby systému syntézy řeči, kde se vstupní dokument (D) zpracuje pomocí výpočetní techniky, kdy zpracování textu (T) se automaticky provede nejméně dvěma metodami zpracování (Mj) k získání výstupu (N) vhodného pro následné další zpracování vedoucí k finální syntéze řeči systémem TTS, přičemž alespoň jedna dílčí metoda zpracování (Mj) je metodou automatické normalizace textu zajišťující identifikaci a normalizaci tokenů specifických pro soudní rozhodnutí. Výstupní reprezentace vstupního textu (N) je zobrazena v SW umožňujícím jeho vizualizaci včetně zobrazení míry důvěry ve výstupy metody pro automatickou normalizaci, přičemž SW umožňuje provádění ručních oprav v automaticky vygenerovaném výstupu (N), přičemž tyto opravy mohou být uchovány pro pozdější použití.

CZ 2023 - 154 A3

Způsob převodu rozhodnutí orgánu veřejné moci z ortografické do fonetické podoby

Oblast techniky

Vynálezem je způsob převodu rozhodnutí orgánu veřejné moci z ortografické do fonetické podoby pro potřeby systému syntézy řeči, kde se vstupní dokument zpracuje pomocí výpočetní techniky.

Dosavadní stav techniky

Současné TTS systémy, tj. Text To Speech - převod textu na mluvenou řeč, již velmi kvalitně generují mluvenou řeč na základě textové předlohy, zejména pokud je tato textová předloha "běžným" textem, tj. např. zprávy, informace o dopravě či počasí, knihy, nebo je z nějaké omezené oblasti (např. dialogový systém nad konkrétním tématem, GPS navigace apod.). TTS systémy lze z pohledu způsobu převodu vstupů (textů) na výstup (audio s řečí) rozdělit na 2 základní skupiny.

První skupinu, tzv. end-to-end systémy, tvoří TTS systémy založené převážně na statistických modelech (často neuronových sítích), jejichž vstupem je na jedné straně ortografická podoba textu a na druhé straně tyto systémy generují přímo výstup ve formě zvukového signálu mluvené řeči. Tyto systémy vyžadují velké množství vstupních dat, která jsou využita pro trénování statistických modelů (neuronových sítí), a fungují jako tzv. black-boxy, tj. je velmi těžké opravovat nějaké konkrétní problémy vyskytující se ve výsledné řeči jako je kupříkladu špatná výslovnost nějakého konkrétního slova apod.). Do těchto systémů lze tedy jen velmi těžko nějak zvenku zasáhnout jinak, než např. úpravou či rozšířením množiny dat pro jejich trénování tak, aby se dosáhlo odstranění daného problému. Bohužel ani tak nelze vždy zaručit, že takový problém bude opravdu odstraněn.

Druhou skupinu tvoří systémy, které provádí zpracování vstupního textu po jednotlivých krocích, tj. zpracování je provedeno zřetězením různých metod zpracování M_i, kde na vstupu jedné metody je výstup jiné metody (nebo výstupy jiných metod). Tyto metody zpracování mohou interně využívat různých přístupů ke zpracování dat, včetně statistických metod a neuronových sítí. Mezi takové metody zpracování může patřit například normalizace textu zahrnující kupříkladu rozvinutí zkratek, číslic či různých symbolů do plné podoby, další metodou může být převod do fonetické podoby F, určení přízvuků či jiných prozodických charakteristik, výpočet doby trvání jednotlivých hlásek, výpočet požadované základní hlasivkové frekvence F_0 či generování posloupnosti A různých typů akustických koeficientů, které pak mohou být dalšími metodami zpracování převedeny na řečový signál R. V případě výskytu nějakého systematického problému ve výsledné řeči R se lze na takový problém zaměřit a jednotlivé metody zpracování modifikovat či měnit jejich parametry tak, aby byl daný problém odstraněn.

Pro tuto druhou skupinu systémů je tedy nutné navrhnout jednotlivé metody zpracování M_i a způsob jejich zřetězení tak, aby výsledný řečový signál R, vygenerovaný pro konkrétní vstupní text T, byl pokud možno co nejlepší. Zároveň je možné systémy z této skupiny snáze přizpůsobit konkrétním specifickým požadavkům nebo třeba specifickému typu vstupních textů, které vyžadují specifické zpracování odlišné od běžného textu. Může se jednat např. o informace o počasí, matematické nebo chemické vzorce, lékařské zprávy, nebo právnické texty. Ze stavu techniky jsou známy kupříkladu dokumenty popisující kupříkladu zpracovávání řečových nahrávek, jako kupříkladu CZ308878, CN109005480, CA 2197734 nebo US 2006173563, přičemž tyto postupy či způsoby zpracování jsou alespoň tvořeny dílčími metodami zpracování. Stav techniky rovněž reprezentuje dokument US9626955B2, ve kterém je předmětem ochrany způsob převodu textu na řeč, kdy na elektronickém zařízení s procesorem a pamětí, která ukládá jeden nebo více programů k provedení převodu procesorem, toto zařízení analyzuje předmětný vstupní dokumentu za účelem identifikace množství textových prvků v dokumentu, které mají být převedeny na řeč, vytvoření mluveného popis kontextu souvisejícího s vytipovanými směrnicemi, které obsahují úryvek kódu s relativním odkazem na paměť, jako je proměnná, adresa URL, obrázek či blok, kdy následně je generován zvuk, který zahrnuje mluvenou formu podmnožiny textu a směrnice. Dílčí stav techniky

rovněž reprezentuje dokument US4692941A jehož předmětem je zařízení a systém převodu textu na řeč v reálném čase.

Mezi metody (M_i) zpracování využívané v těchto systémech patří kupříkladu.:

- Metoda filtrace textu - metoda libovolným známým algoritmem odfiltruje ze vstupního textu T znaky, které se v textu běžně nevyskytují, jako např. HTML tagy, formátovací značky a podobně.

- Metoda normalizace textu - metoda libovolným známým algoritmem převádí například zkratky, číslovky, data, časy, apod. ve vstupním textu T do plně rozvinuté podoby, a to případně včetně správného čísla a pádu dle kontextu, a vytváří tak normalizovaný text N.

- Metoda fonetické transkripce - metoda libovolným známým algoritmem převádí ortografickou podobu vstupního textu T nebo jeho normalizovanou formu N do jeho fonetické podoby F, tj. do posloupnosti jednotlivých fonémů nebo jiných fonetických jednotek

- Metoda určení pozice přízvuků - metoda libovolným známým algoritmem určí ve fonetické podobě F vstupního textu T pozice pro umístění příznaků pro přízvuk.

- Metoda určení prozodických událostí - metoda libovolným známým algoritmem určí ve fonetické podobě F vstupního textu T pozice různých prozodických událostí, např. hranice frází, vět, či jiných prozodických úseků;

- Metoda určení trvání hlásek - metoda libovolným známým algoritmem určí trvání L_j jednotlivých fonémů nebo jiných fonetických jednotek F_j ve fonetické podobě vstupního textu F.

- Metoda určení základní hlasivkové frekvence F_0 - metoda libovolným známým algoritmem určí ve fonetické podobě F vstupního textu T požadovanou amplitudu F_{0i} pro jednotlivé fonémy nebo jiné fonetické jednotky F_j.

- Metoda generování akustických parametrů - metoda libovolným známým algoritmem vygeneruje na základě fonetické podoby F vstupního textu T a na základě dalších již určených charakteristik posloupnost takových akustických parametrů A_k, která dále umožní vygenerování řečového signálu R. Množina typů možných akustických parametrů je odborníkům známa ze stavu techniky, přičemž vlastnosti různých typů akustických parametrů mohou ovlivňovat jejich výběr pro konkrétní použití v konkrétním systému TTS.

- Metoda generování řečového signálu - metoda libovolným známým algoritmem vygeneruje na základě posloupnosti akustických parametrů A_k a na základě dalších již určených charakteristik řečový signál R reprezentující původní vstupní text T.

- Metoda dynamického výběru řečových jednotek - metoda libovolným známým algoritmem vybere na základě fonetické podoby F vstupního textu T a na základě dalších již určených charakteristik z předem připraveného inventáře řečových jednotek l takovou posloupnost řečových jednotek U_l která bude po spojení a případné aplikaci dalších metod zpracování akustického signálu představovat řečový signál R reprezentující původní vstupní text. Řečové jednotky U_l jsou úseky řečového signálu reálné mluvené řeči (délky takových úseků záleží na konkrétní implementaci metody) nahrané jedním mluvčím za specifických podmínek pro potřeby daného systému TTS.

- Metoda závěrečného zpracování - metoda libovolným známým algoritmem finálně

zpracuje řečový signál R tak, aby vyhovoval požadavkům (např. na hlasitost, rychlost, apod.)

Každý TTS systém z této druhé skupiny systémů pak může využívat jen některé z výše uvedených metod zpracování, jejich libovolnou kombinaci, příp. i další libovolné známé, avšak zde neuvedené metody zpracování. Konkrétní implementace metody filtrace textu, metody normalizace textu a metody fonetické transkripce mohou být vybrány na základě typu vstupních textů, v tomto vynálezu v podobě rozhodnutí orgánů veřejné moci. Tyto metody pak mohou také umožňovat jistou flexibilitu a určitou úroveň přizpůsobení se právě konkrétní úloze.

Pro použití v konkrétní specializované oblasti pak nemusí být použití libovolného běžně dostupného TTS systému vhodné, protože tento systém nemusí být vzhledem ke konkrétním implementacím jednotlivých metod schopný korektně zpracovat vstupní psaný text T tak, aby ho výsledný řečový signál R správně reprezentoval.

Jednou z těchto specializovaných oblastí je i oblast převodu soudních či správních rozhodnutí z textové podoby na mluvenou řeč. Důvodem pro pokud možno plně automatický převod právě soudních rozhodnutí do mluvené podoby je skutečnost, že v České republice legislativa požaduje, aby každé soudní rozhodnutí zaznělo v soudní síni. Některé soudní rozhodnutí jsou tak dlouhá, že soudci jsou nuceni několik hodin či dokonce dnů v soudní síni tato rozhodnutí předčítat. Plně automatický nebo alespoň poloautomatický převod soudních rozhodnutí do zvukové podoby by tak těmto soudcům značně ulevil. Zároveň je nezbytné, aby zvuková podoba soudních rozhodnutí naprosto přesně reprezentovala textovou podobu takového rozhodnutí.

V současnosti není znám způsob, který by umožnil takové zpracování textů, zejména soudních rozhodnutí tak, aby výstupem byla fonetická reprezentace umožňující pomocí některých výše uvedených metod, a/nebo i dalších zde neuvedených metod, vygenerování řečového signálu reprezentující tento vstupní text.

Podstata vynálezu

Podstatou vynálezu je způsob převodu textu rozhodnutí orgánů veřejné moci z ortografické do fonetické podoby tak, aby bylo možné textovou podobu takového textu, zejména ve formě soudního rozhodnutí, následně převést do mluvené podoby. Zpracováním se pro podstatu vynálezu rozumí především:

- převod vstupního dokumentu, obsahujícího text soudního rozhodnutí, z různých formátů (čistý text, Microsoft Word, PDF apod.) do podoby, která umožní následné zpracování,
- filtrace nežádoucích znaků a formátovacích značek, které se mohou vyskytovat ve vstupním dokumentu, z nichž některé mohou být dále využity v průběhu dalšího zpracování či v průběhu převodu textu na řeč systémem TTS, např. pro umístění různých prozodických událostí,
- detekce takových úseků vstupního textu (tokenů), které vyžadují zpracování procesem normalizace textu,
- automatická normalizace textu s využitím nově vynalezených metod poskytujících míru důvěry ve výstup těchto metod pro každý jednotlivý token, který byl detekován
- vizualizace zobrazující výstup normalizace textu a míry důvěry ve výstupy normalizace pro jednotlivé tokeny ve specializovaném SW nástroji pro vizualizaci
- umožnění ručních zásahů do procesu normalizace textu pro opravu možných chyb

prostřednictvím specializovaného SW pro zaručení bezchybného zpracování,

- sledování a uchovávání ručních oprav pro možné budoucí vylepšení tohoto procesu zpracování.

Nově vynalezené metody pro automatickou normalizaci textu zajišťují rozepisování či přepis různých tokenů, zapsaných v různých formátech, do plně rozvinuté textové formy a také správného gramatického tvaru, kupříkladu pádu a čísla, který je ovlivněn mimo jiné kontextem.

Tyto metody automatické normalizace textu tedy zajišťují zejména:

a. rozepisování zkratek, specifických pro soudní rozhodnutí, v závislosti na kontextu, a to kupříkladu: č.j. → číslo jednací, o.z. → občanský zákoník, sp. zn. → spisová značka, sb. → sbírka, § → paragraf, tr. zák. → trestní zákoník, a podobně;

b. rozepisování dalších zkratek vyskytujících se i v běžném textu, přičemž je v případě konfliktu potřeba rozhodnout, zda se má zkratka rozepsat dle specifických pravidel (viz bod a), nebo podle běžných pravidel pro normalizaci textu;

c. rozepisování různých bodů struktury logického členění, které mohou být značeny římskými či arabskými číslicemi, nebo velkými či malými písmeny;

d. rozepisování čísel jednacích a čísel spisů, která mohou být zapsána v různém formátu a obsahovat různé znaky;

e. přepis úseků zvýrazněných v textu prokládáním bílými (netisknutelnými) znaky;

f. rozepisování dalších čísel, která mohou být zapsána v různém formátu, například rodných čísel, čísel účtů, registračních značek apod.;

g. rozepisování římských čísel v závislosti na kontextu;

h. rozepisování iniciál (pro následné hláskování při převodu textu na řeč prostřednictvím systému TTS);

i. rozepisování časových údajů, dat apod.

j. rozepisování emailových adres či odkazů na webové stránky;

k. rozepisování různých symbolů v závislosti na kontextu (například závorky);

l. označení neznámých slov, tj. slov, která se nevyskytují v předem definovaném seznamu známých slov (tzv. slovníku);

m. označení hranic prozodických úseků (pro zlepšení kvality následného převodu textu na řeč systémem TTS);

n. označení takových tokenů, u kterých nelze automatickou normalizaci provést.

Vizualizace a ruční zásahy do procesu normalizace textu je zajištěna prostřednictvím specializovaného SW, který umožňuje mj.:

a. vizualizaci původního textu se zobrazením výstupu automatické normalizace textu včetně míry důvěry v tento výstup;

b. na základě míry důvěry ve výstup normalizace textu zvýraznit takové tokeny, u kterých je nutné provést ruční kontrolu správnosti, případně ruční opravu;

c. zaznamenávat všechny ruční opravy pro další zlepšování nově vyvinutých metod automatické normalizace textu;

d. zobrazit umístění prozodických událostí (např. hranice, frází, vět, či jiných prozodických úseků) a jejich ruční úpravu (např. pro zvýšení srozumitelnosti);

e. vytvářet vlastní uživatelský seznam známých slov (tj. slovník) buď jen pro daný dokument nebo i s přesahem do dalších dokumentů, přičemž uživatelské slovníky mohou být v daném dokumentu okamžitě aplikovány pro zlepšení přesnosti normalizace textu;

f. provést převod textu na řeč systémem TTS po jednotlivých odstavcích textu;

g. provést převod textu na řeč systémem TTS pro celý dokument;

Objasnění výkresu

Příkladné provedení navrhovaného řešení je popsáno s odkazem na obr. 1, kde je znázorněno celkové obecné schéma způsobu zpracování převodu textu rozhodnutí orgánu veřejné moci z ortografické do fonetické podoby pro potřeby systému syntézy řeči.

Příklad uáskutečnění vynálezu

Podle navrhovaného vynálezu je proveden způsob převodu soudních rozhodnutí z ortografické do fonetické podoby pro využití v systému syntézy řeči z textu v této oblasti.

Soudci českých soudů, případně jimi určené osoby, musí často veřejně v soudní síni předčítat velmi dlouhá soudní rozhodnutí. Nemohou se tak věnovat jiné činnosti, což může mimo jiné přispívat k neefektivitě soudního systému.

Automatizace vytváření zvukové stopy reprezentující dané soudní rozhodnutí, která by pak mohla být v soudní síni jen přehrána, tak zefektivňuje práci samotných soudců či jimi určených osob. Zvukovou stopu se připravuje na základě daného soudního rozhodnutí s využitím uvedeného vynálezu poloautomatickým způsobem.

Technický prostředek pro poloautomatickou normalizaci textu je ve formě zařízení uzpůsobené k zadání vstupního soudního rozhodnutí, ke zpracování tohoto soudního rozhodnutí níže uvedeným způsobem, a s možností nechat tento výstup dále zpracovat systémem TTS, který vygeneruje výslednou zvukovou stopu.

Při tomto příkladném způsobu zpracování soudních rozhodnutí se každé soudní rozhodnutí zadá do zařízení, dále se využije automatická část, ve které se nově vynalezenými metodami provede detekci tokenů ve vstupním textu soudního rozhodnutí, dále se provede vlastní normalizace tokenů, specifických pro soudní rozhodnutí, kdy zároveň je ke každému normalizovanému tokenů přiřazena míra důvěry ve výstup normalizace. Následně je výsledek této automatické normalizace prezentován určené osobě, která s využitím nově specializovaného SW na daném zařízení provede kontrolu automatické části a případně opraví chyby, které mohly při automatickém zpracování vzniknout. Tyto opravy jsou uloženy v systému pro následné postupné zlepšování vynalezených metod. Nakonec je možné provést vygenerování zvukové stopy systémem TTS.

Zpracování textu T se automaticky provede nejméně dvěma metodami zpracování M_i k získání

výstupu N vhodného pro následné další zpracování vedoucí k finální syntéze řeči systémem TTS, přičemž alespoň jedna dílčí metoda zpracování M_i, je metodou automatické normalizace textu T zajišťující identifikaci a normalizaci tokenů specifických pro soudní rozhodnutí. Alternativně lze tento způsob využít kupříkladu pro rozhodnutí správní.

Průmyslová využitelnost

Vynález lze využít při převodu textu rozhodnutí orgánů veřejné moci z ortografické do fonetické podoby. Jeho využití je předpokládáno zejména při převodu textu soudních rozhodnutí z ortografické do fonetické podoby, dále při převodu textu správních rozhodnutí z ortografické do fonetické podoby nebo rovněž při převodu odborných právnických textů z ortografické do fonetické podoby.

PATENTOVÉ NÁROKY

1. Způsob převodu textu rozhodnutí orgánů veřejné moci z ortografické do fonetické podoby pro potřeby systému syntézy řeči, kde se vstupní dokument (D) zpracuje pomocí výpočetní techniky, **vyznačující se tím**, že zpracování textu (T) se automaticky provede nejméně dvěma metodami zpracování (M_i) k získání výstupu (N) vhodného pro následné další zpracování vedoucí k finální syntéze řeči systémem převodu textu na mluvenou řeč, přičemž alespoň jedna dílčí metoda zpracování (M_i) je metodou automatické normalizace textu pro identifikaci a normalizaci tokenů.

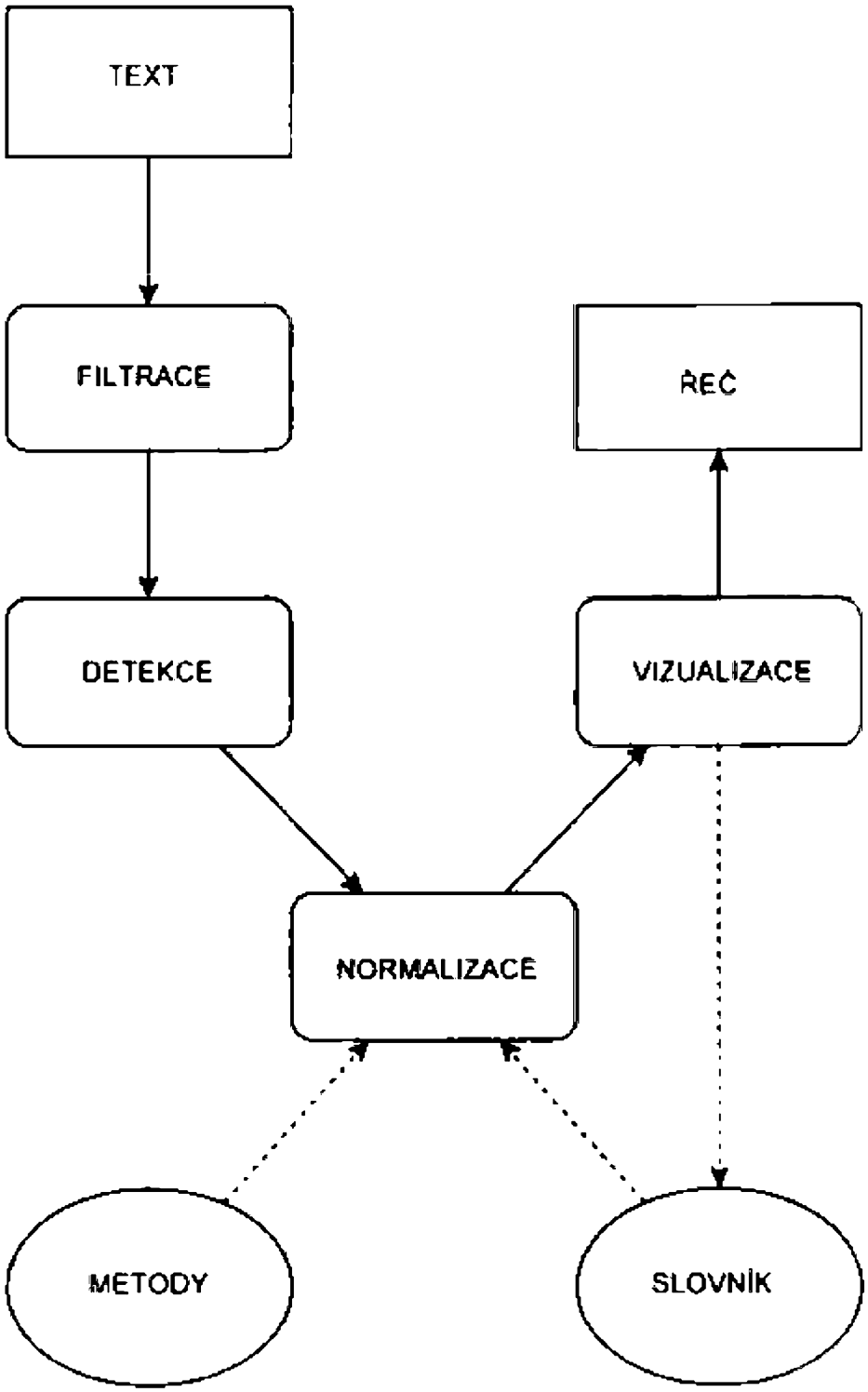
2. Způsob převodu textu z ortografické do fonetické podoby podle nároku 1, **vyznačující se tím,** že výstupní reprezentace vstupního textu (N) je zobrazena v SW umožňujícím jeho vizualizaci včetně zobrazení míry důvěry ve výstupy metody pro automatickou normalizaci.

3. Způsob převodu textu z ortografické do fonetické podoby podle nároku 1 nebo 2, **vyznačující se tím**, že SW je uzpůsoben k provádění ručních oprav v automaticky vygenerovaném výstupu (N).

4. Způsob převodu textu z ortografické do fonetické podoby podle nároku 1, 2 nebo 3, **vyznačující se tím**, že opravy v automaticky vygenerovaném výstupu (N) jsou uloženy pro pozdější použití.

5. Způsob převodu textu z ortografické do fonetické podoby podle nároku 1, 2, 3 nebo 4, **vyznačující se tím**, že výstup reprezentace vstupního textu (N) je zobrazen v SW umožňujícím vytvoření zvukové stopy reprezentující vstupní text (T) libovolným systémem převodu textu na mluvenou řeč.

1 výkres


TEXT
FILTRACE
DETEKCE
NORMALIZACE
METODY
SLOVNÍK
VIZUALIZACE
ŘEČ

Obr. 1