

**Uchazečům**

V Brně dne 30. 10. 2012  
Č.j.: VZ 2/2012

**Věc: „VZ: CEITEC – Software k vyhodnocení výsledků pulzní gelové elektroforézy“, dodatečné informace č.1 – odpověď na dotaz uchazeče**

Číslo veřejné zakázky: **VZ 2/2012**

Název: **CEITEC – Software k vyhodnocení výsledků pulzní gelové elektroforézy**

**Uchazeč** podal dne 24. 10. 2012 v 19:34 na emailovou adresu kontaktní osoby zadavatele [vystaveloval@vfu.cz](mailto:vystaveloval@vfu.cz) (Ing. Lenka Vystavělová, Veterinární a farmaceutická univerzita Brno, Palackého tř. 1/3, Brno 612 42) – níže uvedený **dotaz**:

Dobrý den,

rád bych Vás poprosil o doplňující informace k Vámi vypsané veřejné zakázce "CEITEC – Software k vyhodnocení výsledků pulzní gelové elektroforézy".

Chtěli bychom se zeptat, zda jsou k dané zakázce k dispozici nějaké dodatečné informace.

Dál by mě zajímalo několik věcí k jednotlivým položkám zadání (tučně):

*Obecné požadavky:*

*objektově orientované relační multi-uživatelské databázové rozhraní, s až 500 informačními poli pro každou položku.*

***Jaká konkrétně informační pole máte na mysli a co je zde myšleno položkou? Je to jednotlivý experiment nebo něco jiného?***

*ukládání informací v podobě textu, čísel či dat.*

*propojování položek s více experimenty.*

***Máte na mysli propojení ve smyslu, že jednomu projektu bude patřit více experimentů či jakou konkrétně máte představu?***

*vyhledávání záznamů mezi databázemi v závislosti na informačních polích, experimentech a obsahu.*

*multidatabázový systém – každá databáze může obsahovat kombinaci různých experimentů*

*možnost exportu do XML souborů a importu XML souborů do databáze*

*Vytváření tzv. fingerprintů (denzitometrických profilů píků či bandů (proužků)):*

*1. zpracování a normalizace obrazu*

*možnost importu fingerprintů z bitmapových obrazů, denzitogramů a chromatogramů*

***Co je zde míněno fingerprintem? Proč a jak má aplikace pracovat s chromatogramy a v jakém budou formátu? Bitmapové obrazy mají představovat pouze denzitogramy nebo ještě něco jiného?***

*pre-editace a úprava obrazu*

**Bude se jednat pouze o 1D elektroforézu nebo chcete i 2D nebo nD elektroforézu? Náročnost vývoje software toto jistě může značně ovlivnit.**

*automatická detekce drah („lanes“) u všech typů gelů*

*nastavitelné parametry pro jednotlivé dráhy*

**Máte konkrétní představu o parametrech, které je třeba nastavovat?**

*automatické a manuální zarovnání („alignment“), rozpoznávání vzorců s využitím externích referenčních vzorců a/nebo interních referenčních bandů*

**V jakém formátu budou vstupy externích referenčních vzorců? Interní referenční bandy budou ve stejném formátu jako běžný experiment?**

*normalizace bitmapových obrazů s indikací spolehlivosti a možných chybných zarovnaní, odstraňování skvrn, nastavitelné vyhlazování křivek a odečítání pozadí.*

**Jak konkrétně si představujete indikaci spolehlivosti?**

## 2. kvantifikace

*algoritmy pro vyhledávání bandů s nastavitelnou citlivostí*

*možnost najít a označit nejisté bandy/píky*

*kvantifikace molekulárních velikostí bandů*

*přesné vyjádření množství/koncentrací proteinů či nukleových kyselin v závislosti na známých kalibračních píkách*

*komparativní kvantifikace bandů/píků mezi skupinami vzorů*

*vyhledávání diskriminativních bandů/píků mezi vybranými skupinami vzorů,*

**Můžete lépe specifikovat pojem skupiny vzorů, co konkrétně znamená vzor a jaký je vstupní formát vzoru?**

*vyhledávání unikátních a běžných bandů/píků*

**Jaký je rozdíl mezi unikátním a běžným bandem?**

*možnost přidávat či editovat bandy přímo při srovnávání*

**Můžete prosím specifikovat zadání? Přidáním bandu myslíte nakreslit band na místo kde nebyl?**

2

*Shluková (clusterová) analýza a vytváření dendrogramů:*

### 1. clusterová analýza

*umožnění srovnání až 20 000 databázových záznamů, s možností využití různých metod*

*(Unweighted pair-grouping (UPGMA), complete linkage (furthest neighbor), single linkage (nearest neighbor), Ward, Centroid, Median, Neighbor Joining, Bio-Neighbor*

*Joining, NeighborNet clustering)*

**Můžete specifikovat zda se jedná o externí databázi nebo máte na mysli spíše o záznamy v databázi, které budou vytvořeny používáním zadaného programu?**

*nastavitelná optimalizace a hranice tolerance pro vzory bandů („banding patterns“)*

**Máte na mysli implementaci konkrétního (publikovaného) algoritmu?**

*statistická determinace nejvhodnějších hranic tolerance pro vzory bandů*

**Máte na mysli implementaci konkrétního (publikovaného) algoritmu?**

*výpočet a vizualizace spolehlivosti shlukování pro všechny algoritmy a skupiny dat*  
**Jak je definována spolehlivost shlukování?**

*kompozitní shluková analýza – možnost spojení různých datových setů do jedné konsenzuální shlukové analýzy, výpočet globální podobnosti.*

*2. interpretace*

*kombinovatelné zobrazení obrazů, sekvencí, normalizovaných vzorů, maticemi podobnosti, s tříděním podle dendrogramu*

***Můžete konkretizovat? Jaké sekvence chcete zobrazovat? Normalizovaným vzorem máte na mysli diagram 1D elektroforézy po preprocessingu daným programem? Jaké matice podobnosti chcete použít a jak jsou definované (máte nějakou konkrétní preferenci použitého algoritmu pro porovnávání dvou vzorů)? Třídění podle dendrogramu je třídění v pořadí vzorů v jakém je vykreslený dendrogram?***

*indikace statistické chyby shlukování*

***Podle mých znalostí nelze standardní chybu pro shlukovou analýzu spočítat, protože se jedná o metodu hledání podobnosti předmětů analýzy podle konkrétního algoritmu, nejedná se tedy o statistický test a jako taková neexistuje žádná standardní statistická chyba shlukování. Můžete tedy prosím specifikovat konkrétní algoritmus, který chcete použít pro výpočet takové chyby?***

*přímá interakce mezi databází a dendrogramem*

*přidávání či odebírání záznamů z existujících shlukových analýz bez přepočítávání celé analýzy*

*možnost ukládání srovnání na disk*

*vytváření 2D a 3D grafů, kontingenčních tabulek*

*Součástí předmětu plnění je také:*

*dodávka a doprava předmětu plnění na místo plnění veřejné zakázky;*

*odborná instalace softwaru a uvedení do plně funkčního a provozuschopného stavu;*

*zaškolení obsluhy - určeného pracovníka kupujícího*

*provedení všech předepsaných zkoušek, revizí, seřízení, vystavení nutných protokolů,*

*atestů případně jiných právních nebo technických dokladů, kterými bude prokázáno*

*dosažení předepsané kvality a předepsaných technických parametrů předmětu plnění;*

*předání technické dokumentace s přesným popisem předmětu plnění v českém jazyce,*

*dokumentace bude zástupci kupujícího předána nejpozději při předání předmětu plnění*

Děkuji.

S pozdravem

**Zadavatel podává tuto odpověď:**

**Odpověď k dotazu č.1:**

**Dotaz uchazeče:** „Chtěli bychom se zeptat, zda jsou k dané zakázce k dispozici nějaké dodatečné informace.“

**Odpověď zadavatele:** Zadavatel k zakázce doposud, tj. k datu 30.10.2012, nepodával žádné dodatečné informace. Případné dodatečné informace budou vždy uveřejněny na webových stránkách zadavatele na stejném místě jako zadávací dokumentace, tj. <http://www.vfu.cz/uredni-deska/verejne-zakazky/verejne-zakazky-maleho-rozsahu/index.html>, tak jak je uvedeno ve Výzvě k podání nabídek.

## **Odpověď k dotazu č.2:**

**Dotaz uchazeče:** „Dál by mě zajímalo několik věcí k jednotlivým položkám zadání (tučně).“ Jednotlivé dotazy uchazeče jsou v následujícím textu vyznačeny tučně a označeny jako „Dotaz:“.

**Odpověď zadavatele:** Jednotlivé odpovědi zadavatele jsou uvedeny v následujícím textu vždy za příslušným dotazem uchazeče, odpovědi jsou vyznačeny tučně červeným písmem a označeny jako „Odpověď:“.

- Obecné požadavky:

- *objektově orientované relační multi-uživatelské databázové rozhraní, s až 500 informačními poli pro každou položku.*

**Dotaz: Jaká konkrétně informační pole máte na mysli a co je zde myšleno položkou? Je to jednotlivý experiment nebo něco jiného?**

**Odpověď: „Položkou“ je myšlen jednotlivý vzorek či záznam v databázi. „Informační pole“ je v tomto smyslu obecné označení jednoho konkrétního parametru (vlastnosti), jehož různé hodnoty chce uživatel jednotlivým položkám přiřazovat. Tj. například barva, datum sběru, druh, lokalita... jednoduše řečeno možnost uložit až 500 různých informací přiřazených k jednotlivé databázové položce.**

- *ukládání informací v podobě textu, čísel či dat.*

- *propojování položek s více experimenty.*

**Dotaz: Máte na mysli propojení ve smyslu, že jednomu projektu bude patřit více experimentů či jakou konkrétně máte představu?**

**Odpověď: Ano, k jednomu projektu bude patřit více experimentů. „Experimentem“ je myšlen soubor dat z určitého biologického pokusu či testu, tj. např. restriční profil získaný elektroforézou genomové DNA štěpené jedním restričním enzymem. Protože tentýž vzorek může být štěpen i jiným enzymem a získá se tak jiný restriční profil, avšak vztahující se k témuž vzorku, je potřeba možnost propojit každé položku (vzorek) s více experimenty.**

- *vyhledávání záznamů mezi databázemi v závislosti na informačních polích, experimentech a obsahu.*

- *multidatabázový systém – každá databáze může obsahovat kombinaci různých experimentů*

- *možnost exportu do XML souborů a importu XML souborů do databáze*

- Vytváření tzv. fingerprintů (denzitometrických profilů píků či bandů (proužků)):

### 1. zpracování a normalizace obrazu

- *možnost importu fingerprintů z bitmapových obrazů, denzitogramů a chromatogramů*

**Dotaz: Co je zde míněno fingerprintem? Proč a jak má aplikace pracovat s chromatogramy a v jakém budou formátu? Bitmapové obrazy mají představovat pouze denzitogramy nebo ještě něco jiného?**

**Odpověď: Fingerprintem je míněn denzitometrický profil píků či bandů. Aplikace nebude pracovat přímo s chromatogramy (tj. např. výstupy z analýzy plynovou chromatografií či HPLC, ale s příslušnými fingerprinty). Vstupní formát souborů je 8-bitový TIFF, JPEG, GIF či PNG. Bitmapové obrazy představují denzitogramy (elektroforegramy apod.)**

- *pre-editace a úprava obrazu*

Dotaz: **Bude se jednat pouze o 1D elektroforézu nebo chcete i 2D nebo nD elektroforézu? Náročnost vývoje software toto jistě může značně ovlivnit.**

Odpověď: **Jedná se o 1D elektroforézu.**

- *automatická detekce drah („lanes“) u všech typů gelů*

- *nastavitelné parametry pro jednotlivé dráhy*

Dotaz: **Máte konkrétní představu o parametrech, které je třeba nastavovat?**

Odpověď: **Tloušťka dráhy a definice uzlů, kde jsou dráhy zakřivené či ohnuté.**

- *automatické a manuální zarovnání („alignment“), rozpoznávání vzorců s využitím externích referenčních vzorců a/nebo interních referenčních bandů*

Dotaz: **V jakém formátu budou vstupy externích referenčních vzorců? Interní referenční bandy budou ve stejném formátu jako běžný experiment?**

Odpověď: **„Externím referenčním vzorcem“ je myšlen například standardní vzorek nanesený na různých pozicích téhož zpracovávaného experimentu (např. elektroforetického gelu). Ten je pak jako reference využíván i při zpracování dalších denzitogramů z následných experimentů téhož druhu. Vzorek je tedy definován uvnitř programu a stejně jako interní referenční bandy (interní ve smyslu vyskytující se pouze v právě zpracovávaném experimentu) je ve stejném formátu jako běžný experiment.**

- *normalizace bitmapových obrazů s indikací spolehlivosti a možných chybných zarovnání, odstraňování skvrn, nastavitelné vyhlazování křivek a odečítání pozadí.*

Dotaz: **Jak konkrétně si představujete indikaci spolehlivosti?**

Odpověď: **Např. jako barevné vyjádření lokálních odchylek posunu při normalizaci vzhledem k celkovému posunu referenčního vzoru vzhledem k referenčním pozicím.**

## 2. kvantifikace

- *algoritmy pro vyhledávání bandů s nastavitelnou citlivostí*

- *možnost najít a označit nejisté bandy/píky*

- *kvantifikace molekulárních velikostí bandů*

- *přesné vyjádření množství/koncentrací proteinů či nukleových kyselin v závislosti na známých kalibračních píkách*

- *komparativní kvantifikace bandů/píků mezi skupinami vzorů*

- *vyhledávání diskriminativních bandů/píků mezi vybranými skupinami vzorů,*

Dotaz: **Můžete lépe specifikovat pojem skupiny vzorů, co konkrétně znamená vzor a jaký je vstupní formát vzoru?**

Odpověď: **Vzorem je myšlen denzitometrický profil (fingerprint) jednoho vzorku. Skupinou vzorů pak např. několik vybraných profilů pocházejících z téhož experimentu. Příklad: elektroforetický gel s dvanácti vzorky; každý jeden vzorek vykazuje určitý vzor (pattern) bandů; když budu chtít srovnat tyto patterns mezi čtyřmi určitými vzorky na tomtéž gelu, pak tyto čtyři patterns tvoří momentální skupinu vzorů. Každý vzor je tedy definován v předchozím kroku, při zpracování a normalizaci obrazu a při případném manuálním označení bandů.**

- vyhledávání unikátních a běžných bandů/píků

Dotaz: **Jaký je rozdíl mezi unikátním a běžným bandem?**

Odpověď: **Unikátní – vyskytující se vzácně či v jediném vzoru; běžný – vyskytující se častěji či v každém vzoru**

- možnost přidávat či editovat bandy přímo při srovnávání

Dotaz: **Můžete prosím specifikovat zadání? Přidáním bandu myslíte nakreslit band na místo kde nebyl?**

Odpověď: **V podstatě ano: „řít“ programu „v tomto místě je band“, i když tam nebyl programem automaticky nalezen.**

- Shluková (clusterová) analýza a vytváření dendrogramů:

### 1. clusterová analýza

- umožnění srovnání až 20 000 databázových záznamů, s možností využití různých metod (Unweighted pair-grouping (UPGMA), complete linkage (furthest neighbor), single linkage (nearest neighbor), Ward, Centroid, Median, Neighbor Joining, Bio-Neighbor Joining, NeighborNet clustering)

Dotaz: **Můžete specifikovat zda se jedná o externí databázi nebo máte na mysli spíše o záznamy v databázi, které budou vytvořeny používáním zadaného programu?**

Odpověď: **Platí druhá varianta – záznamy vytvořené v zadaném programu.**

- nastavitelná optimalizace a hranice tolerance pro vzory bandů („banding patterns“)

Dotaz: **Máte na mysli implementaci konkrétního (publikovaného) algoritmu?**

Odpověď: **Ne.**

- statistická determinace nevhodnějších hranic tolerance pro vzory bandů

Dotaz: **Máte na mysli implementaci konkrétního (publikovaného) algoritmu?**

Odpověď: **Ne.**

- výpočet a vizualizace spolehlivosti shlukování pro všechny algoritmy a skupiny dat

Dotaz: **Jak je definována spolehlivost shlukování?**

Odpověď: **Myšlena je kongruence shlukování vzorků jednotlivými algoritmy. Spolehlivost je nepřesný termín.**

- kompozitní shluková analýza – možnost spojení různých datových setů do jedné konsenzuální shlukové analýzy, výpočet globální podobnosti.

### 2. interpretace

- kombinovatelné zobrazení obrazů, sekvencí, normalizovaných vzorů, maticemi podobnosti, s tříděním podle dendrogramu

Dotaz: **Můžete konkretizovat? Jaké sekvence chcete zobrazovat? Normalizovaným vzorem máte na mysli diagram 1D elektroforézy po preprocessingu daným programem? Jaké matice podobnosti chcete použít a jak jsou definované (máte nějakou konkrétní preferenci použitého algoritmu pro porovnávání dvou vzorů)? Třídění podle dendrogramu je třídění v pořadí vzorů v jakém je vykreslený dendrogram?**

Odpověď: **Zobrazení sekvencí není požadováno, zůstalo omylem jako relikt.**

**Ad normalizovaný vzor – ano.**

**Ad matice: požadované jsou koeficienty Jaccard, Dice, Jeffrey's X, Ochiai**

**Ad třídění podle dendrogramu – ano.**

*- indikace statistické chyby shlukování*

Dotaz: **Podle mých znalostí nelze standardní chybu pro shlukovou analýzu spočítat, protože se jedná o metodu hledání podobností předmětů analýzy podle konkrétního algoritmu, nejedná se tedy o statistický test a jako taková neexistuje žádná standardní statistická chyba shlukování. Můžete tedy prosím specifikovat konkrétní algoritmus, který chcete použít pro výpočet takové chyby?**

Odpověď: **Jedná se o požadavek na statistické vyjádření signifikance zjištěných clusterů. Samotný dendrogram poskytuje informaci o vytvořených skupinách, ale ne už o spolehlivosti příslušného větvení na každé úrovni větvení.**

**Nejjednodušší indikací signifikance větví jsou průměrné podobnosti jednotlivých větví v dendrogramu. Směrodatná odchylka každé větve je získána rekonstrukcí hodnot podobností vzorků v dané větvi dendrogramu a srovnání těchto hodnot s původními hodnotami podobností. Standardní odchylka odvozených (rekonstruovaných) hodnot a původních hodnot je pak mírou spolehlivosti a vnitřní konzistence dané větve.**

**Jinou možností je kofenetická korelace, která vypočítává korelaci mezi podobnostmi odvozenými z dendrogramu a maticí podobnosti.**

*- přímá interakce mezi databází a dendrogramem*

*- přidávání či odebrání záznamů z existujících shlukových analýz bez přepočítávání celé analýzy*

*- možnost ukládání srovnání na disk*

*- vytváření 2D a 3D grafů, kontingenčních tabulek*

Součástí předmětu plnění je také:

*- dodávka a doprava předmětu plnění na místo plnění veřejné zakázky;*

*- odborná instalace softwaru a uvedení do plně funkčního a provozuschopného stavu;*

*- zaškolení obsluhy - určeného pracovníka kupujícího*

*- provedení všech předepsaných zkoušek, revizí, seřízení, vystavení nutných protokolů, atestů případně jiných právních nebo technických dokladů, kterými bude prokázáno dosažení předepsané kvality a předepsaných technických parametrů předmětu plnění;*

*- předání technické dokumentace s přesným popisem předmětu plnění v českém jazyce, dokumentace bude zástupci kupujícího předána nejpozději při předání předmětu plnění*



## Zadavatel dále oznamuje uchazečům, že:

V souvislosti s výše uvedenými dotazy uchazeče a podávanými dodatečnými informacemi **zadavatel prodlužuje lhůtu pro podání nabídek.**

**Nový konec lhůty pro podání nabídek** je stanoven na **12.11.2012, 13:00 hod včetně.**

Otevírání obálek s nabídkami proběhne dne: **12.11.2012 v 13:01 hod,** a to v sídle kanceláří zadavatele (tj. Veterinární a farmaceutická univerzita Brno, Rektorát - Oddělení veřejných zakázek - zasedací místnost, I. poschodí, dveře č. 211).

**Ostatní podmínky uvedené ve výzvě k podání nabídek a zadávací dokumentaci zůstávají zachovány.**

**Uchazeči, kteří již podali nabídku, jsou oprávněni vzít svoji nabídku zpět.**

  
Mgr. Daniela Němcová  
kvestorka VFU Brno

Vyřizuje: Ing. Lenka Vystavělová  
[vystaveloval@vfu.cz](mailto:vystaveloval@vfu.cz), 54156 2012