Техничко решење: Технологија аутоматског препознавања говора на српском и њему сродним језицима	
Аутори техничког решења:	Владо Делић, Никша Јаковљевић, Радован Обрадовић, Дарко Пекар, Драгиша Мишковић, Драган Кнежевић
Назив техничког решења:	Технологија аутоматског препознавања говора на српском и њему сродним језицима
Подтип техничког решења:	Нова технологија на међународном нивоу (М81)
Реализатори:	ФТН и АлфаНум доо из Новог Сада

Пројекат и период реализације техничког решења:

2005-2008: Развој говорних технологија за српски језик и примена у Телекому (Пројекат технолошког развоја МНТР: ТР-1644А)

2008-2010: Говорна комуникација човек-машина (Пројекат технолошког развоја МНТР: ТР-1644А)

За кога је техничко решење рађено:

Највеће интересовање је иницијално показао Телеком, али и остали оператери фиксне и мобилне телефоније, како у Србији, тако и у окружењу. Међутим, примена овог решења далеко превазилази овај скуп корисника.

Потенцијални корисници су:

- Телеком оператери
- Све фирме које поседују сопствене позивне центре
- Правна лица која се баве пружањем услуга путем говорних аутомата
- Компаније које имају савремене телефонске централе и желе да их унапреде увођењем препознавања говора
- Систем интегратори који се баве испоруком централа, говорних аутомата и позивних центара
- Физичка и правна лица која имају потребе да у некој својој апликацији уведу препознавање говора на командном нивоу (паметне куће, складишта, тржни центри, управљање радом рачунара, итд.).

Ко техничко решење користи (прихватио-примењује):

Техничко решење је нашло примену у више говорних аутомата имплементираних на територији бивше Југославије.

Примери су:

- Централна депозитарна агенција, Подгорица
- Поштанаска штедионица, Београд

- Говорни аутомат за слепе - Контакт, Србија
- Говорни аутомат за слепе - Телевечерњак, Хрватска
- Телефонски сервис за упознавање - Састанак, Србија
- Аутобуска станица Крагујевац
- Аутопревозник Ласта, Србија
- Аутобуска станица Бања Лука
- Компанија Фонлидер, Београд
- Компанија Сага, Београд
- Компанија Логос, Загреб
- Компанија Сагена, Загреб.

Како су резултати верификовани (од стране кога тела):

1) Ово техничко решење базирано је на више техничких и развојних решења која су пријављивана као резултати на технолошким и иновационим пројектима МНТР у периоду 2005-2010. године и прво је такво техничко решење у ширем региону.
2) Писано мишљење два рецензента-експерта из области техничког решења.
3) Департман за енергетику, електронику и телекомуникације ФТН-а издао је Уверење о признавању техничког решења на основу извештаја Комисије за техничка решења ДЕЕТ које потврђује да испуњава све услове да буде признато као нови производ на међународном нивоу (М81), у складу са документом МНТР "*ПРАВИЛНИК о поступку и начину вредновања, и квантитативном исказивању научноистраживачких резултата истраживача*" од 21.03.2008. године.

На који начин се користи (кратак опис):

Техничко решење је имплементирано као софтверски модул за препознавање речи и фраза из малих и средњих речника на српском и језицима са сличном фонетиком (хрватски, бошњачки). Основна функционалност модула је да препозна која реч или фраза из задатог скупа је изговорена у одређеном аудио фајлу. Модул такође враћа и информацију о поузданости, као и податак о томе да ли је изговорена реч изван речника.

Овај модул користе крајње апликације (говорни аутомати, позивни центри, системи за управљање и сл.) и на тај начин обезбеђују говорни интерфејс према кориснику. Као интерфејс између апликације и модула за препознавање може се користити један од неколико имплеметираних интерфејса: Microsoft SAPI, IP, C++ библиотека, COM.

Опис техничког решења: Технологија аутоматског препознавања говора на српском и другим јужнословенским језицима

Систем за аутоматско препознавање говора на српском језику представља скуп софтверских алата који имају за циљ конверзију говорног сигнала у одговарајући текст. Развијан је у оквиру неколико пројеката Министарства, прилагођавањем савремених, општеприхваћених алгоритама из области обраде сигнала, препознавања облика и машинског учења специфичностима српског језика. Његов развој пратио је већи број научних и стручних радова као и развојних и техничких решења, који су пријављивани као резултати на пројектима МНТР у периоду од 2005-2010 године.

Област на коју се техничко решење односи:

Техничко решење припада области информационих, односно информационо-комуникационих технологија (ICT) и представља једну од основних говорних технологија.

Проблем који се техничким решењем решава:

Систем за аутоматско препознавање говора престаљa једну од основних компоненти која је неопходна у развоју говорних технологија као што су: говорни интерфејси, системи за претраживање мултимедијалних (аудио) садржаја, говорни дијалошки системи, системи за аутоматско превођење говора и сл.

Говорни интерфејси постепено добијају своје место, поред стандардних улазно излазних уређаја (тастатура, миш, екрани осетљиви на додир, графичко окружење). Тренутно доминантан, графички кориснички интерфејс, са системом менија, захтева велики број операција и за једноставне задатке, што је превазиђено увођењем скраћеница на тастатури (*keyboard shortcut*), а које могу бити замењене (замењују се) гласовним командама. Мобилни телефони добијају све већу процесорску снагу и могу да реализују све већи број различитих задатака, а због својих димензија стандардни улазни уређаји нису погодни, стога се говорни интерфејс јавља као разуман избор. У ситуацијама када су и очи и руке заузети, нпр. у току вожње, говор је најбезбеднији начин комуникације.

Развој широкопојасног Интернета и ниска цена уређаја за смештање и чување података омогућили су доступност информација у форми аудио и видео записа. Пошто је већина претраживача прилагођена за претрагу текстуалних датотека, ове мултимедиалне садржаје је потребно претворити у одговарајући текст, а конверзију аудио садржаја врше системи за аутоматско препознавање говора.

За човека говор представља најприроднији вид комуникације па је и говорни дијалошки систем човеку најближи вид интеракције са машином. Препознавање

говора представља први степен на основу ког систем добија информације о намерама и стању корисника.

У системима за аутоматско превођење говора, први степен представља систем за препознавање говора, који говорни сигнал конвертује у одговарајући текст, који се потом анализира и преводи у текст на неком другом тзв. циљном језику, који се потом синтетизује.

Стање решености тог проблема у свету:

Интензиван развој на пољу препознавања говора је започео већ 70 година прошлог века. Ови иницијални кораци су поставили теоријске основе већине савремених система за препознавање говора. У већини система акустичка варијабилност говорног сигнала је моделована применом скривених Марковљевих модела у комбинацији са Гаусовим мешавинама или вештачким неуралним мрежама односно *support vector machines*, док је језик моделован било помоћу граматичких правила или *n*-грама. Обука модела се реализује углавном помоћу техника заснованих на критеријуму максималне веродостојности, али се користе и технике дискриминативне обуке као што су критеријум максималне међуинформације, минималне грешке класфикације и минималне грешке на нивоу речи односно фонема. Препознавање говора се врши углавном применом Витербијевог алгоритма, али постоје системи који користе стек (stack) алгоритам. Да би се смањила нежељена варијабилност говорног сигнала користе се различите технике нормализације и адаптације, као што су нормализација средњом вредношћу и варијансом кепстрала, нормализација дужине вокалног тракта, нормализација на основу хистограма, адаптација на говорника и канал применом максимум апостериори критеријума или линеарне регресије базиране на критеријуму максималне веродостојности.

У зависности од тежине задатка препознавања, перформансе система изражене преко грешке на нивоу речи, варира од једног до четрдесетак процената (што је број речи већи, говор спонтанији, зашумљенији и измешан са другим звуцима то је грешка на нивоу речи већа). Ни један од постојећих комерцијалних и некомерцијалних система не подржава препознавање говора на српском (хрватском однсон бошњачком) језику.

Објашњење суштине техничког решења и детаљан опис са карактеристикама, укључујући и пратеће илустрације и техничке цртеже (техничке карактеристике):

Као и код већине савремених система за препознавање говора и код овог система моделовање говора на акустичком нивоу је реализовано помоћу скривених Марковљевих модела (*Hidden Markov Model* – HMM) и Гаусових мешавина (*Gaussian Mixture Model – GMM*). За разлику од већине сличних некомерцијалних система Гаусове расподеле су моделоване помоћу пуних коваријансних матрица уместо дијагоналних. На овај начин је постигнуто нешто боље покривање акустичког простора

са мањим бројем мешавина у односу на случај када се користе искључиво дијагоналне коваријансне матрице.

Основна јединица моделовања је глас зависан од контекста, тзв. трифон. Број стања којим се моделује један трифон варира од један до шест, а сразмеран је просечном трајању фонема чију реализацију представља трифон. Избором гласова зависних од контекста за јединице моделовања, повећана је тачност модела, али је број опсервација који је на располагању за обуку смањен. Да би се унапредила ефикасност естимације параметара трифона врши се повезивање *HMM* стања. У оквиру овог система имплементирана су два решења. Прво решење је ефикасније у случају малог броја опсервација, а базира се на лингвистичким знањима о сличности између фонема и претпоставци да фонем који се налази у левом контексту има више утицаја на почетак гласа, а фонем у десном контексту на крај гласа. Друго решење је реализовано по узору на кластеровање на бази стања (*tree based clustering*), где се при повезивању поред фонетске блискости контекста узима у обзир акустичка блискост стања. За разлику од оригиналног алгоритма, имплементирано кластеровање врши се над опсервацијама уместо над моделима, а варијације опсервација које су последица различитих канала и говорника сведене на минимум.

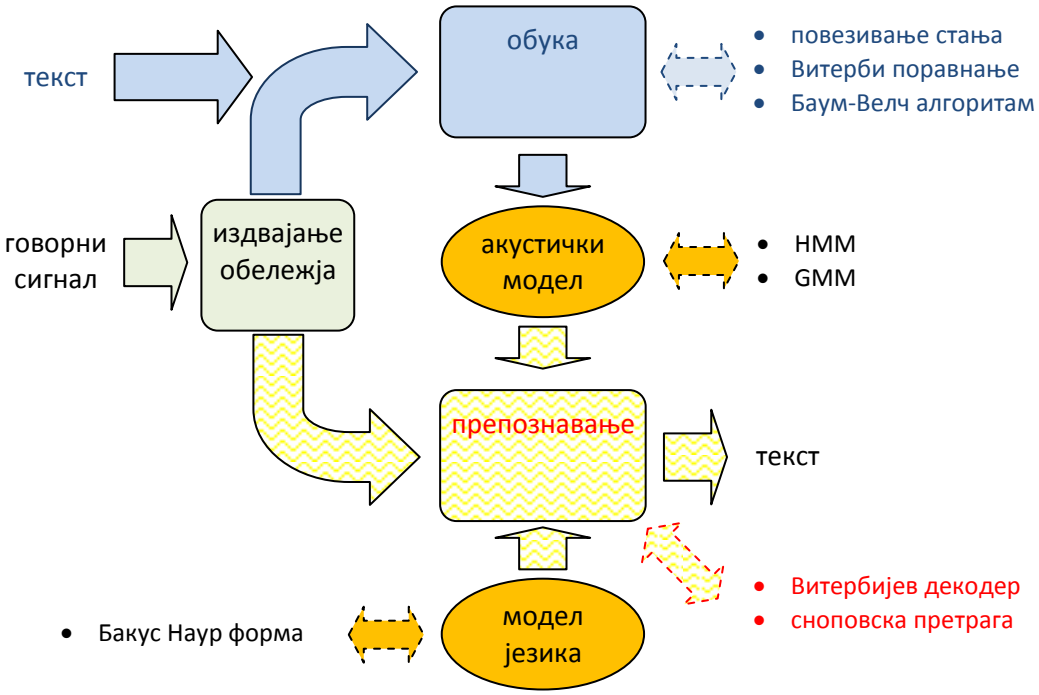
Естимација параметара акустичких модела је базирана на критеријуму максималне веродостојности. Подржана је прерасподела опсервација између стања на основу Витербијевог (Viterbi) алгоритма код ког једна опсервација припада само једном стању и Баум-Велчовог (Baum-Welch) алгоритма код ког једна опсервација припада неколицини стања са различитим вероватноћама. Прерасподела опсервација између смеша које чине једно HMM стање врши се помоћу *k-means* алгоритма, Гаусовог класификатора или EM алгоритма.

Број Гаусових расподела по стању се одређује хеуристички, тј. број Гаусових расподела се постепено повећава све док вредност просечне веродостојности на верификацијоном скупу опсервација не почне да пада. Након што се одреди број Гаусових расподела, опсервације из верификационог скупа се укључују у естимацију осталих параметара модела (средње вредности, коваријансне матрице, тежине мешавина).

Декодовање се реализује применом Витербијевог алгоритма. Да би се убрзала претрага подржано је такозвано поткресивање (*pruning*), при којем се путање које имају знатно мању вероватноћу од вероватноће до тог тренутка највероватније путање одбацују. Путање које преживе поткресивање чине тзв. сноп, тако да се ова метода назива још и претрага у снопу (*beam search*). У случају да сноп чини мали број путања, претрага је брза али може доћи до одбацивања најбоље секвенце. Са друге стране у случају да сноп чини велики број путања, препознавање је успорено. Стога је потребно

наћи компромис, у овом систему он је реализован адаптивним постављањем прага за поткресивање.

Модел језика се задаје ручно помоћу граматичких правила у Бакус Наур (Backus Naur) форми. Пошто је систем намењен за препознавање скупа од неколико стотина речи, то је више него довољно.



Како је реализован и где се примењује, односно које су могућности примене (техничке могућности):

Систем је реализован у програмском језику C++. Кôд је организован у 3 библиотеке и то: библиотеку опште намене, библиотеку за проточну обраду сигнала и библиотеку функција везаних за обуку и тестирање система за препознавање говора. Библиотека опште намене обезбеђује основне класе и функције за манипулацију текстуалним фајловима, лабелама, архивирањем података и сл. Библиотека за проточну обраду сигнала садржи класе које реализују основне операције које постоје у дигиталној обради сигнала (филтрирање, естимација спектра, естимација енергије итд.), али и операције које су везане за издвајање обелжја говорног сигнала (мел-фреквенцијске кепстралне коефицијенте, банке филтара, линеарне предиктивне коефицијенте итд.). У оквиру библиотека функције везаних за обуку и тестирање система за препознавање говора имплементирани су алгоритми који су описани у претходном одељку.

Тренутна имплементација због манипулација са звучном картицом и аудио фајловима везана је за Windows OS.

Техничке могућности:

- AlfaNumASR подржава препознавање до 1000 речи у једном тренутку. Теоретски, могуће је ићи и преко те границе, али постепено долази до нарушавања поузданости и брзине рада модула.
- На речнику од 200 речи систем прави грешку испод 2% (*word error rate*).
- У питању је фонетски препознавач, што значи да се скуп речи за препознавање дефинише у текстуалном облику, а модели речи се праве "у лету", преко већ споменутих трифона.
- Систем враћа и податак о степену поузданости препознавања. У случају ниске поузданости, на апликацији је да предузме одговарајућу стратегију (верификација одговора, понављање питања, нуђење помоћи, пребацивање на оператера).
- Модул подржава приступ више корисника истовремено. У зависности од величине граматике и брзине рачунара, могуће је вршити и до неколико десетина препознавања истовремено, без кумулативног кашњења.
- Оптимизован је за рад са фајловима телефонског квалитета. Ово је последица тога што је говорна база која служи за обуку снимана преко телефона. С обзиром на прве примене, то је и била намера. Обуком на другој говорној бази лако се добија систем прилагођен другачијим условима.
- Подржано је више интерфејса, што обезбеђује лако уграђивање овог модула у произвољну апликацију.

Докази (прилози):

- Писано мишљење два рецензента-експерта из области техничког решења
- Фактуре према фирмама које су ово техничко решење уградиле у своје апликације
- Постојање више говорних аутомата на јавно доступним телефонским бројевима (Контакт - 0700/200500, Састанак - 021/4873366, Аутобуска станица Крагујевац - 0902/034034, АлфаНум демо - 021/4873350, ЦДА Подгорица - +382 20 444900)

Штампано у јулу 2010.

УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА
Трг Доситеја Обрадовића 6, 21000 Нови Сад, Република Србија
Деканат: 021 6350-413; 021 450-810; Централа: 021 485 2000
Рачуноводство: 021 458-220; Студентска служба: 021 6350-763
Телефакс: 021 458-133; e-mail: ftndean@uns.ac.rs
ИНТЕГРИСАНИ СИСТЕМ МЕНАЏМЕНТА СЕРТИФИКОВАН ОД:

Наш број:
Ваш број:
Датум: 2010-07-02

ИЗВОД ИЗ ЗАПИСНИКА

Наставно-научног већа Факултета техничких наука у Новом Саду, на 9. редовној седници одржаној дана 16.06.2010. године, донело је следећу одлуку:

-непотребно изостављено-

Тачка 13. Питања научноистраживачког рада и међународне сарадње

У циљу доношења одлуке о прихватању техничког решења под називом:

ТЕХНОЛОГИЈА АУТОМАТСКОГ ПРЕПОЗНАВАЊА ГОВОРА НА СРПСКОМ И НА ЊЕМУ СРОДНИМ ЈЕЗИЦИМА

Аутори: Владо Делић, Никша Јаковљевић, Радован Обрадовић, Дарко Пекар, Драгиша Мишковић, Драган Кнежевић

именују се рецензенти:

1. др Слободан Јовичић, ванредни професор ЕТФ у Београду
2. др Зоран Перић, ванредни професор ЕФ у Нишу

-непотребно изостављено-

Записник водила: Јасмина Димић, дипл. правник

Тачност података оверава: Секретар Иван Нешковић, дипл. правник

Декан Проф. др Илија Босић

РЕЦЕНЗИЈА ТЕХНИЧКОГ РЕШЕЊА

Подаци о техничком решењу:

Назив техничког решења:	Технологија аутоматског препознавања говора на српском и њему сродним језицима
Аутори техничког решења:	Владо Делић, Никша Јаковљевић, Радован Обрадовић, Дарко Пекар, Драгиша Мишковић, Драган Кнежевић
Реализатори:	ФТН и АлфаНум доо из Новог Сада
Подтип техничког решења:	Нова технологија на међународном нивоу (М81)

Подаци о рецензенту:

Име, презиме и звање:	Др Зоран Перић, ванредни професор
Ужа научна област за коју је изабран у звање, датум избора у звање и назив факултета:	изабран у звање ванредног професора 15.10.2005. на Електронском факултету у Нишу за у.н.о. Телекомуникације
Установа где је запослен:	Електронски факултет у Нишу

Стручно мишљење рецензента:

"Технологија аутоматског препознавања говора на српском и њему сродним језицима" представља **нову технологију уведену у производњу на међународном нивоу (М81)** у смислу Правилника МНТР за техничка решења од 21.03.2008. године. Образложење:

- Овај систем (говорна технологија) за аутоматско препознавање говора на српском и сродним језицима представља скуп софтверских алата који имају за циљ конверзију говорног сигнала у одговарајући текст.
- Постоје слична решења за друге језике, међутим ово техничко решење је достигло потребан ниво и нашло комерцијалне примене у посебним производима и сервисима у неколико држава у региону.
- Оригинални научни методи развијени у различитим фазама рада на пројекту МНТР, чији је један део верификован публикацијом у међународним часописима, чине теоријску основу развијене технологије. Практична реализација технологије и њено тестирање је изведено на сопственој опреми.

У Нишу , 09.07.2010. године.

Проф. др Зоран Перић

РЕЦЕНЗИЈА ТЕХНИЧКОГ РЕШЕЊА

Подаци о техничком решењу:

Назив техничког решења:	Технологија аутоматског препознавања говора на српском и њему сродним језицима
Аутори техничког решења:	Владо Делић, Никша Јаковљевић, Радован Обрадовић, Дарко Пекар, Драгиша Мишковић, Драган Кнежевић
Реализатори:	ФТН и АлфаНум доо из Новог Сада
Подтип техничког решења:	Нова технологија на међународном нивоу (М81)

Подаци о рецензенту:

Име, презиме и звање:	Др Слободан Јовичић, ванредни професор
Ужа научна област за коју је изабран у звање, датум избора у звање и назив факултета:	Реизабран у звање ванредног професора 18.09.2007. на ЕТФ у Београду за у.н.о. Говорна комуникација
Установа где је запослен:	ЕТФ Београд и ЦУЖА Београд

Стручно мишљење рецензента:

"Технологија аутоматског препознавања говора на српском и њему сродним језицима" представља **нову технологију уведену у производњу на међународном нивоу (М81)** у смислу Правилника МНТР за техничка решења од 21.03.2008. године. Образложење:

- Ова говорна технологија представља скуп поступака или алгоритама за аутоматско препознавање говора на српском и језицима са сличним фонетским инвентаром.
- Имплементирана је као скуп софтверских алата који имају за циљ конверзију говорног сигнала у одговарајући текст. Постоје слична решења за друге језике, а ово техничко решење је још увек једино које је достигло потребан ниво и нашло комерцијалне примене у посебним производима и сервисима у неколико држава у региону.
- Развијена је коришћењем научних метода претежно на сопственој опреми и без коришћења туђе патентне или лиценцне документације. Као резултат развојних активности произашао је већи број радова, међу којима је и рад у SCI часопису.

У Београду, 07.07.2010. године.



Проф. др Слободан Јовичић

УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА
Трг Доситеја Обрадовића 6, 21000 Нови Сад, Република Србија
Деканат: 021 6350-413; 021 450-810; Централа: 021 485 2000
Рачуноводство: 021 458-220; Студентска служба: 021 6350-763
Телефакс: 021 458-133; e-mail: ftndean@uns.ac.rs
ИНТЕГРИСАНИ СИСТЕМ МЕНАЏМЕНТА СЕРТИФИКОВАН ОД:

Наш број:
Ваш број:
Датум: 15.07.2010.

ИЗВОД ИЗ ЗАПИСНИКА

Наставно-научног већа Факултета техничких наука у Новом Саду, на 11. редовној седници одржаној дана 14.07.2010. године, донело је следећу одлуку:

-непотребно изостављено-

Тачка 13. Питања научноистраживачког рада и међународне сарадње

На основу мишљења рецензената усваја се техничко решење М81 под називом:

ТЕХНОЛОГИЈА АУТОМАТСКОГ ПРЕПОЗНАВАЊА ГОВОРА НА СРПСКОМ И НА ЊЕМУ СРОДНИМ ЈЕЗИЦИМА

Аутори: Владо Делић, Никша Јаковљевић, Радован Обрадовић, Дарко Пекар, Драгиша Мишковић, Драган Кнежевић

-непотребно изостављено-

Записник водила:
Јасмина Димић, дипл. правник

Тачност података оверава:
Секретар
Иван Нешковић, дипл. правник

Декан
Проф. др Илија Ћосић

Pl. 29.12.

(93)

ALFANUM
GOVORNE TEHNOLOGIJE

Trg Dositeja Obradovića 6
Novi Sad, Srbija
021/4750080
PIB: 102948614
Mat.br. 08790949
Tekući račun: 340-11002697-66
Šifra delatnosti: 72400
IBAN: RS35340000879094900064
SWIFT:GIBARS22
SWIFT CODE FOR CORRESPONDENT BANK:GIBAATWW

Račun br. 080/08

Kupac

Naziv	Fonlider d.o.o.
Adresa	Sime Igumanova 2-4
Mesto	11000 Beograd
P.I.B	100206806

Datum:	30.dec.08
Mesto:	Novi Sad
DPO:	30.dec.08

Br.art.	Opis	Količina	Jed. mere	Cena bez PDV (din.)	Rabat	Por. Osnova (din.)	Stopa PDV	PDV (din.)	Ukupno (din.)
	Licenca za Alfanum ASR	5	kom.	21.958,56		109.792,80	18%	19.762,70	129.555,50
2	ALF H4 STD USB	1	kom.	2.652,00		2.652,00	18%	477,36	3.129,36

Ukupno bez PDV-a	112.444,80
Poreska osnovica	112.444,80
PDV iznos	20.240,06
Za uplatu	132.684,86

Slovima: sto trideset dve hiljade šesto osamdeset četiri dinara i 86/100

Rok plaćanja :

Ovlašćeno lice

ČENOM ODGOVORNOŠĆU DRUŠTVO SA OGRANIČENOM ODGOVORNOŠĆU ALFANUM GOVORNE TEHNOLOGIJE NOVI SAD SRBIJA I CRNA GORA

Pl. 06.08.

46

ALFANUM
GOVORNE TEHNOLOGIJE

Trg Dositeja Obradovića 6
Novi Sad, Srbija
021/4750080
PIB: 102948614
Mat.br. 08790949
Registarski broj: 002890219
Šifra delatnosti: 72400
Tekući račun: 335-6188-21
IBAN : RS35335007010003218895, Metals banka AD
SWIFT : MBSORS22

Račun br. 041/08

Kupac

Naziv	Centralna Depozitarna Agencija AD
Adresa	Novaka Miloševa bb
Mesto	81000 Podgorica, Crna Gora
P.I.B	

Datum	22.jul.08
Mesto:	Novi Sad
Datum PDO:	22.jul.08

Br.art.	Opis	Količina	Jed. mere	Cena bez PDV (€)	Rabat	Por. Osnova (€)	Stopa PDV	PDV (€)	Ukupno (€)
1	Softver za primanje poziva	1	kom.	2.970,00		2.970,00			2.970,00

Ukupno bez PDV-a	2.970,00
Poreska osnovica	2.970,00
PDV iznos	0,00
Za uplatu	2.970,00

Slovima: dve hiljade devetsto sedamdeset eura

Rok plaćanja : 15 dana

Ovlašćeno lice

DRUŠTVO SA OGRANIČENOM ODGOVORNOŠĆU ALFANUM GOVORNE TEHNOLOGIJE NOVI SAD SRBIJA I CRNA GORA

Usluga namenjena izvozu. PDV se ne zaračunava po članu 24, zakona, stav 1, tačka 2, Sl. glasnik 84/04