

L'intertextualité dans les publications scientifiques

Cyril Labbé, Dominique Labbé

▶ **To cite this version:**

Cyril Labbé, Dominique Labbé. L'intertextualité dans les publications scientifiques. L'Auteur scientifique, Jun 2013, Grenoble, France. hal-00840769

HAL Id: hal-00840769
https://hal.science/hal-00840769

Submitted on 3 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L’Auteur scientifique

Séminaire
Laboratoire Lidilem (Linguistique et Didactique des langues)
Université Stendhal (Grenoble III)

28 juin 2013

Cyril Labbé
Laboratoire d’Informatique de Grenoble - Université Joseph Fourier
(cyril.labbe@imag.fr)

Dominique Labbé
Laboratoire PACTE (CNRS - Institut d’Etudes Politiques de Grenoble)
(dominique.labbe@iep.grenoble.fr)

L'intertextualité dans les publications scientifiques

Résumé
La base de données bibliographiques de l’IEEE contient un certain nombre de duplications avérées avec indication des originaux copiés. Ce corpus est utilisé pour tester une méthode d’attribution d’auteur. La combinaison de la distance intertextuelle avec la fenêtre glissante et diverses techniques de classification permet d’identifier ces duplications avec un risque d’erreur très faible. Cette expérience montre également que plusieurs facteurs brouillent l’identité de l’auteur scientifique, notamment des collectifs de chercheurs à géométrie variable et une forte dose d’intertextualité acceptée voire recherchée.

Abstract
The IEEE bibliographic data base contains a number of proven duplications with indication of the original(s) paper(s) copied. This corpus is used to test an authorship attribution method. The intertextual distance, combined with the sliding window and with various classification techniques, identifies these duplications with a very low risk of error. This experience also shows that several factors blur the identity of the scientific author, including the variable geometry of collective research and the high levels of intertextuality accepted and, sometimes, desired.

Sommaire

I. Qui a écrit ce texte scientifique ? 3

Traitements préalables 4

Distance intertextuelle 4

Limites de validité de l'estimation de la distance 5

Facteurs déterminant la distance 6

Valeurs anormales 7

L'identité fluctuante de l'auteur scientifique 8

Une forte intertextualité assumée 10

II. L'intertextualité dissimulée **12**

Duplication sans référence ni permission 12

Détection des proximités entre originaux et duplications 13

Trois catégories principales de duplications 16

Détection des passages dupliqués : la fenêtre glissante 18

Validation et détermination des paramètres optimaux 19

Conclusions 21

Bibliographie 22

Annexe 1. Le corpus des duplications et des originaux signalés par l'IEEE 23

Annexe 2. Les 53 distances anormalement faibles dans le corpus O ($\alpha = 0.3\%$) 60

Annexe 3. Classification des duplications (D) en fonction de leur(s) plus proche(s) voisin(s) originaux (O) avec leurs distances respectives (D(d,o)) 66

Annexe 4. Un exemple de détection des passages dupliqués avec la fenêtre glissante 72

S’interroger sur l’auteur d’un texte scientifique peut paraître paradoxal pour deux raisons.

Premièrement, la principale règle de la communication scientifique est l’impersonnalité obtenue par l’effacement des empreintes qui marquent habituellement la présence de l’auteur dans son texte, tout particulièrement les pronoms personnels – au premier rang le « je » et le « vous » - la plupart des marques d’énonciation, les figures rhétoriques, etc.

Deuxièmement, les textes scientifiques contemporains portant sur un même sujet utilisent non seulement la même terminologie et les mêmes concepts mais ils se rapportent obligatoirement à des travaux antérieurs. Dans beaucoup de communications, il existe une section "related works", passant en revue la "littérature" existante ; des extraits de certains travaux importants sont cités ou résumés.

Ce phénomène est communément baptisé "intertextualité" ou *présence d’un texte dans un autre texte*. L’intertextualité est à la base même de la communication scientifique mais elle doit s’afficher et respecter des règles assez précises (usage des guillemets – ou d’autres embrayeurs équivalents -, noms des auteurs du document cité et références bibliographiques).

Il paraît donc difficile de retrouver l’auteur d’un texte scientifique avec les techniques utilisées pour les textes littéraires (Love 2002 ; Koppel & Al 2009 ; Stamatatos 2011 ; Savoy 2012). Pourtant, nous allons montrer que cette identification est possible en replaçant la question de l’auteur dans une problématique plus générale : la classification d’un grand nombre d’individus, ici des textes (labbé & Labbé 2011).

Une première expérience montrera que la méthode proposée rend possible l’identification de l’auteur. En effet, on verra que les textes issus d’une ou de plusieurs plumes identiques partagent une proportion plus importante de leur vocabulaire que des textes écrits par des auteurs différents. Ceci permettra d’identifier les cas d’intertextualité cachée.

I. Qui a écrit ce texte scientifique ?

Les textes utilisés pour ces expériences sont des communications scientifiques portant sur l’informatique et l’électronique, extraits de la base de données bibliographiques de l’IEEE (Annexe 1). L’IEEE (Institute of Electrical and Electronic Engineers) est avec l’ACM (Association for Computing Machinery) la principale association pour l’informatique. Sa base bibliographique (payante) serait la plus vaste dans le domaine de l’informatique, de l’électronique et des sciences connexes (plus de 3 millions de références au printemps 2011 selon l’IEEE).

La première expérience porte sur 259 textes composant le corpus "O" – pour "originaux" dont les principales caractéristiques sont décrites dans le tableau ci-dessous.

Tableau I. Principales caractéristiques du corpus O (259 textes)

	Plus ancien	Plus récent	Médiane	Mode
Année	1997	2011	2005	2007
	Plus petite	Plus longue	Moyenne	95%
Longueurs (mots)	1 207	75 670	5 476	1 712 – 12 068

Outre le fait d'être des publications scientifiques dans le champ de l'informatique (et sciences connexes), ces textes ont une caractéristique commune : leur paternité n'est pas contestée… Ce corpus n'a pas été constitué selon les méthodes classiques de l'échantillonnage au hasard. Cependant, comme on le verra plus bas, la sélection des textes a été effectuée par un grand nombre d'opérateurs indépendants les uns des autres. Une vérification empirique permettra de constater que cette sélection aboutit à des caractéristiques comparables à celles d'un échantillon sélectionné selon les règles classiques.

Traitements préalables

Les fichiers pdf sont convertis en textes bruts à l'aide du programme "pdftotxt" (programme unix libre version 3.01 pour windows). Cette opération est le "maillon faible" de la procédure car certains textes sont mal reconnus et des améliorations seraient nécessaires. Au cours de l'opération, les figures, les graphiques, les formules, les chiffres disparaissent. En revanche, les titres et les légendes sont conservés. Les bibliographies sont retirées pour des raisons exposées par ailleurs (Labbé & Labbé 2012a).

Dans une première phase, on a testé des variantes à cette conversion automatique : conservation des chiffres et des symboles mathématiques ; élimination des entêtes et pieds de page, laboratoires et adresses des auteurs, etc. Il s'agissait d'observer les éventuelles distorsions introduites par un traitement entièrement automatique, seul possible sur de très vastes populations de textes.

Les textes sont ensuite segmentés en mots (word tokens) suivant la procédure implémentée dans le programme "Oxford Concordance Program" (Hockey & Martin, 1998) – le plus utilisé pour les textes en anglais. Puis les vocables (word types) sont classés par ordre alphabétique et l'on compte leurs occurrences (nombre de fois qu'ils apparaissent dans le texte). Enfin, on mesure les distances entre tous les couples possibles de textes.

Distance intertextuelle

La méthode a été exposée dans Labbé & Labbé 2011 (voir aussi : Labbé & Labbé 2003 ; 2006). Elle consiste à superposer deux textes (A et B) et à compter le nombre de mots différents. Soit :

• N_A et N_B : le nombre de mots (word-tokens) respectivement dans A et B, cad la **longueur** de ces deux textes ;

• V_A et V_B : le nombre de vocables (word-types) respectivement dans A et B, cad le **vocabulaire** de ces deux textes ;

• F_{iA} et F_{iB} : le nombre des occurrences (effectifs) d'un vocable *i* respectivement dans A et B ;

Dans le cas de deux textes de même longueur ($N_A = N_B$) : $|F_{iA} - F_{iB}|$: la différence absolue (sans signe) entre les effectifs d'un vocable dans A et B ; $D_{(A,B)}$: la distance intertextuelle entre A et B est la somme des différences absolues entre les effectifs de tous les vocables de A et B ($V_{(A,B)}$) :

$$D_{(A,B)} = \sum_{i \in (A,B)}^{V_{(A,B)}} \left| F_{iA} - F_{iB} \right| \quad avec\ N_A = N_B \quad (1)$$

L'indice de la distance (ou distance relative) :

$$D_{rel(A,B)} = \frac{\sum_{i \in (A,B)}^{V_{(A,B)}} \left| F_{iA} - F_{iB} \right|}{N_A + N_B} \qquad (2)$$

Dans le cas de deux textes de longueurs inégales ($N_A < N_B$), la distance est estimée en "réduisant" B à la longueur de A :

• $U = \frac{N_A}{N_B}$ est le coefficient utilisé pour réduire B en B' ($N_{B'} = N_A$).

• $E_{iA(u)} = F_{iB}.U$ est l'effectif théorique dans B' d'un vocable *i* ($i \in B$).

Dans la formule (1) ci-dessus, l'effectif de chaque vocables de B est remplacé par cet effectif théorique dans B' :

$$D_{(A,B')} = \sum_{i \in (A,B)}^{V_{(A,B)}} \left| F_{iA} - E_{iA(u)} \right|$$

La formule (2) devient :

$$D_{rel(A,B')} = \frac{\sum_{i \in (A,B)}^{V_{(A,B)}} \left| F_{iA} - E_{iA(u)} \right|}{N_A + N_{B'}} \qquad (3)$$

Cette dernière formule peut se définir ainsi : *estimation du nombre de mots différents que compteraient A et B si B avait la même longueur que A.*

Cet indice présente les trois propriétés attendues d'une distance (dans un espace euclidien) : identité, symétrie, inégalité triangulaire. Il varie uniformément entre 0 – le même vocabulaire est utilisé dans les deux textes avec les mêmes fréquences – et 1 (aucun mot en commun). Un indice de 0,5 signifie que les deux textes partagent 50% des mots, c'est-à-dire la moitié de leur contenu.

Limites de validité de l'estimation de la distance

De nombreuses expériences ont permis de déterminer empiriquement que la bonne qualité des résultats obtenus avec (3) est suspendue à trois conditions :

- les textes doivent être suffisamment longs (au minimum 1000 mots), sinon l'indice est instable et ses valeurs sont toujours très élevées. Dans le corpus O, un texte comportait moins de 1000 mots et a donc été sorti de l'expérience en cours ;

- pour les textes brefs (moins de 3000 mots), l'indice présente parfois une certaine instabilité et ses valeurs peuvent être plus élevées. Cet inconvénient est d'autant plus sensible que les textes sont plus courts. Dans le corpus O, 58 textes – soit 22,5% du total - ont une longueur inférieure

à 3 000 mots ; 10% ont une longueur inférieure à 2 250 mots. Les résultats obtenus sur les textes de ce dernier déciles seront donc examinés à part ;

- la longueur des textes comparés ne doit pas être trop différente. Dans tous les cas, pour les textes en anglais, le rapport entre la plus petite et la plus grande longueur doit être inférieur à 1/7 (Labbé D. 2007). La dernière case du tableau 1 indique que 95% des longueurs sont comprises entre 1712 et 12068 mots (ce qui respecte ce ratio de 1 :7). Les 7 textes (soit 2,5%) les plus courts ont des longueurs comprises entre 1 207 et 1 712 mots ; les 7 textes les plus longs (2,5%) ont des longueurs supérieures à 13 000 mots. Les cinq plus longs sont d'un type un peu particulier (mémoires et thèses). Le raisonnement sera donc valide pour 95% des textes ; les 5% restants faisant l'objet d'un examen à part.

Facteurs déterminant la distance

Ces expériences ont également montré que, dans les limites ci-dessus, l'indice est principalement influencé par 4 facteurs. Par ordre décroissant d'importance : le genre, l'auteur, le thème et l'époque. Le genre – c'est-à-dire le moyen de communication utilisé – est toujours le facteur le plus lourd car il comporte toute une série de règles particulières qui s'ajoutent à celles de la langue. Vient ensuite l'auteur. Autrement dit, deux textes contemporains – dans un même genre – sont toujours séparés par une distance plus petite quand ils sont écrits par un même auteur que par deux auteurs différents et ceci quel que soit le thème.

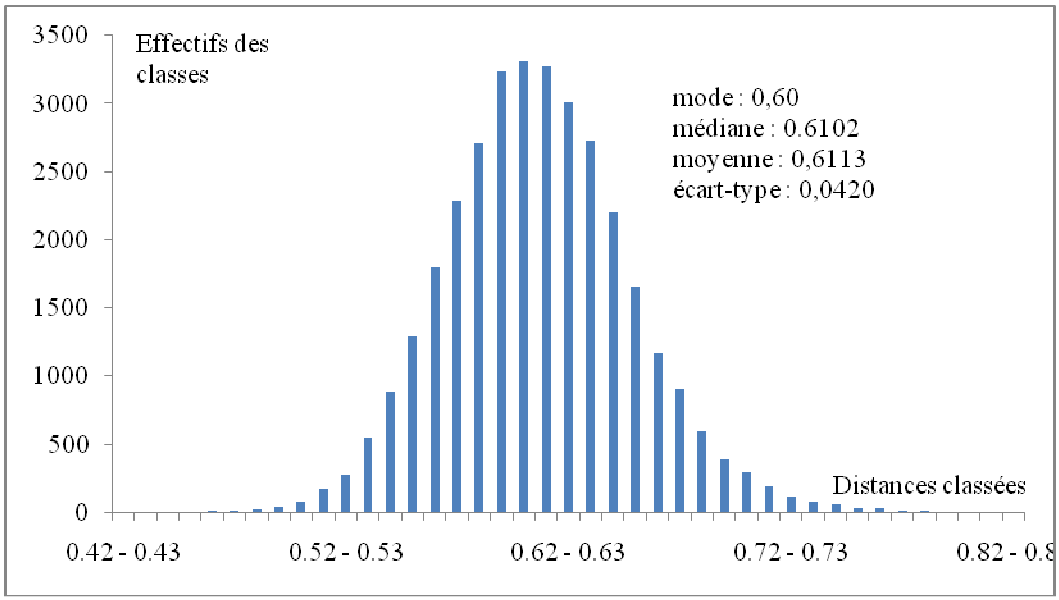
Pour le corpus O, deux de ces quatre facteurs sont neutralisés. Premièrement, les textes appartiennent tous au même genre (littérature scientifique sur l'informatique et l'électronique). Deuxièmement, ils sont tous contemporains (première ligne du tableau 1). A part une dizaine de textes publiés à la fin des années 1990, ils sont tous parus au début des années 2000.

Il ne reste donc que deux facteurs : l'auteur et le thème. Dans les textes littéraires, la correspondance, le discours politique, la presse… la variable auteur l'emporte sur le thème. En est-il de même dans la littérature scientifique ?

Le corpus O permet de tester empiriquement cette question. En effet, quelques auteurs – ou groupes d'auteurs – ont plusieurs textes. La distance séparant ces textes de mêmes auteurs est-elle systématiquement plus faible que celle séparant deux textes d'auteurs différents, même quand le sujet de ces textes est identique ?

Pour répondre à cette question, on calcule d'abord les distances intertextuelles séparant les 256 textes (soit 33 411 couples différents) et l'on range ces distances par ordre croissant dans des classes égales d'intervalle 0,01 (Tableau 2 ci-dessous).

Tableau 2 Histogramme des distances dans le corpus O classées par ordre croissant (intervalle de classe : 0.01)



La forme en cloche du graphique et la confusion des trois valeurs centrales indiquent une distribution gaussienne telle qu'on l'observe sur des échantillons aléatoires, ce qui signifie que le corpus O présente les caractéristiques d'un tel échantillonnage. De plus, la dispersion des valeurs autour de la moyenne est faible : elle est mesurée par l'écart type (noté σ) et par le coefficient de variation relative : les deux tiers des valeurs de la série sont distribuées autour de la moyenne ± 6,9%. Cette faible dispersion indique une population relativement homogène.

Valeurs anormales

Une distribution de ce genre comporte nécessairement quelques valeurs "anormales" (sinon elle ne serait pas "aléatoire"). En fonction du seuil d'erreur accepté (α), on peut définir des intervalles de confiance (tableau 3 ci-dessous).

Tableau 3 Bornes de l'intervalle de confiance en fonction du seuil retenu

	Borne inférieure	Borne supérieure
α = 5%	0,5290	0,6935
α = 1%	0,5039	0,7187
α = 0.3%	0,4854	0,7372

Par exemple, dans le cadre d'une distribution gaussienne on s'attend à ce que 99,7% de la population soit comprise dans l'intervalle $\{\bar{x} \pm 3\sigma\}$. Autrement dit, dans le cas présent, il devrait y avoir 50 distances (0,15% du total) inférieures à la moyenne diminuée de 3 écarts

types (soit 0,485). En fait, il y en a 53, ce qui montre à nouveau la distribution gaussienne des distances. Ces 53 couples "anormalement" proches sont présentés dans le tableau en annexe 2.

D'après le modèle présenté ci-dessus, on s'attend à ce que les distances les plus faibles soient observées entre textes de mêmes auteurs – y compris avec des différences de thèmes – puis entre les textes d'auteurs différents traitant un même sujet, à la même date, etc.

Mais le corpus O contient-il plusieurs textes signés par des mêmes auteurs ? On remarquera que cette situation semble peu probable avec une "base de sondage" de cette ampleur (256 individus tirés aléatoirement parmi plus de 3 millions)… Pourtant le corpus O contient une vingtaine de cas de ce genre, la plupart avec deux textes mais un groupe d'auteurs en a trois et un autre cinq.

Par exemple, la première ligne du tableau en annexe indique que les O0054 et O0055 sont en fait le même texte, par les mêmes auteurs (distance : 0,159) à deux stades différents. Il a d'abord été présenté dans une conférence avec actes[1] puis publié à nouveau dans une revue[2]. En deuxième rang et troisième rang, on trouve des auteurs identiques pour deux textes - portant sur le même sujet - présentés la même année dans deux conférences différentes avec une forte proportion de texte en commun (distance 0.230 pour le deuxième rang et 0,2446 pour le troisième rang). En quatrième rang (distance 0.296), les deux mêmes auteurs ont présenté leurs travaux dans une version courte (4 953 mots) puis longue (14 107 mots)[3]. Etc.

Cependant, l'hypothèse d'un facteur auteur prépondérant ne semble pas totalement vérifiée. En effet, il y a 32 couples possibles avec les mêmes auteurs. On s'attendrait donc à ce que les 32 premières lignes du tableau soient occupées par ces 32 couples. Ce n'est le cas que jusqu'à la ligne 17…

Les couples des 18^{e} et 22^{e} rangs concernent un texte (00260) dont la longueur (11 916 mots) excède la limite supérieure du ratio de 1 à 7 évoqué ci-dessus. En revanche, ce n'est pas le cas pour les couples de rang 20 et 24 et suivants qui portent sur les mêmes sujets mais sont signés par des auteurs différents. L'examen de ces "échecs" apparents fait ressortir les deux caractéristiques singulières de la publication scientifique, sujet de cette journée d'étude : l'identité fluctuante de l'auteur des textes scientifiques ; une intertextualité assumée, voire recherchée.

L'identité fluctuante de l'auteur scientifique

La plupart des papiers scientifiques sont signés par plusieurs auteurs et ces collaborations sont souvent à "géométrie variable". Dans le corpus O, ces variations concernent plus d'une quinzaine de couples de textes. A titre d'exemple, considérons d'abord le groupe d'auteurs qui a cinq textes (Stephen D. J. McArthur et autres auteurs) dont les articles portent sur des sujets identiques ou relativement proches. On attend 10 couples de valeurs anormalement faibles (distances intra) et en tous cas, toutes inférieures à celles séparant l'un de ces 5 textes à l'un des 251 autres textes constituant le corpus O (même lorsque ce texte "inter" porte sur le même sujet). Le tableau 4 ci-dessous donne les résultats avec les rangs et les distances.

[1] Thomas Weng, Yi Zhu & Chung-Kuan Cheng. Digital Design and Programmable Logic Boards: Do Students Actually Learn More? *38th ASEE/IEEE Frontiers in Education Conference*, October 22–25, 2008, Saratoga Springs, NY.

[2] Yi Zhu, Thomas Weng & Chung-Kuan Cheng. Enhancing Learning Effectiveness in Digital Design Courses Through the Use of Programmable Logic Boards. IEE Transactions on Education. Vol 52, n° 1, February 2009.

[3] La distance est ici sous-estimée du fait de la longueur du second qui excède la limité supérieure indiquée ci-dessus.

Tableau 4 Distances les plus faibles concernant les cinq textes signés par S. D. J. McArthur et autres auteurs (O0018, O0034, O0035 ; O00124 ; O0161)

Rang	Texte A	Texte B	Distance(A,B)
9	O0034	O0035	0,3368
12	O0018	O0035	0,3535
20	O0018	O0034	0,4077
25	O0124	O0161	0,4272
28	O0018	O0124	0,4351
30	O0018	O0161	0,4397
34	O0034	O0124	0,4602
38	O0035	O0124	0,4629
51*	**O0023**	**O0035**	**0,4800**
59	**O0023**	**O0034**	**0,4868**
68	O0034	O0161	0,4900
94	**O0096**	**O0124**	**0,4985**
140	O0035	O0161	0,5046

* Pour auteurs et titres voir annexes 1 et 2.

Huit des dix couples sont bien séparés par des distances inférieures à l'intervalle de confiance (avec un seuil de 1%).

Cependant, deux des distances "intra" (rang 68 et surtout 140) sont supérieures à ce seuil et à bon nombre de couples formés par des publications signés par des auteurs différents. L'explication essentielle serait la suivante : les auteurs de ce groupe de textes ne sont pas exactement les mêmes suivant les dates.

- O0034 et 0035 sont parus en 2007, signés par : S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi[4].

- O00161 a été publié en 2004 ; signé par : V. M. Catterson & S. D. J. McArthur (appartenant tous deux à l'Université de Glasgow).

- O0018 est paru en 2005 signé par : Victoria M. Catterson, Euan M. Davidson & Stephen D. J. McArthur (appartenant tous les trois à la même université) ;

- O00124 a été publié en 2004, signé par Stephen D. J. McArthur, Scott M. Strachan & Gordon Jahn (appartenant tous trois à la même université).

Autrement dit, au centre de ces 5 papiers, on trouve bien les travaux d'une même équipe de recherche – dans un temps assez resserré - mais cette équipe voit sa composition changer au fil du temps ; de plus pour deux de ces papiers, elle a noué une collaboration avec 4 chercheurs appartenant à trois équipes différentes (Grèce, Etats-Unis et Japon). Si ces différents auteurs ont effectivement participé à l'écriture des papiers, il est normal qu'ils soient séparés par des distances plus importantes que s'ils avaient eu toujours les mêmes rédacteurs.

[4] S. D. J. McArthur, E.M. Davidson & V. M. Catterson : University of Strathlcyde, Glasgow ; A. L. Dimeas & N. D. Hatziargyriou : University of Athens ; F. Ponci : University of South Carolina ; T. Funabashi : Meidensha Corporation (Tokyo).

Le groupe d'auteurs ayant trois textes dans le corpus O présente les mêmes difficultés :

- O0028 et signé par S. Khan, K. F. Li, E. G. Manning, M. M. Akbar.

- O0080 : Md Mostofa Akbar, Eric G. Manning, Gholamali C. Shoja & Shahadat Khan.

- O0081 : Md Mostofa Akbar, M. Sohel Rahman, M. Kaykobad, E.G. Manning & G.C. Shoja.

Seuls M. M. Akbar et E. G. Manning sont signataires des trois papiers. Entre les 028 et 081 – qui sont aussi les plus distants (d = .434) – cinq des sept auteurs sont différents.

A partir du 20e rang, tous les couples de textes remarquables présentent des caractéristiques de ce genre : ordre des auteurs différents, signataires qui disparaissent ou apparaissent d'un papier à l'autre... Cela met en valeur une difficulté qui n'existe pas, du moins à ce point, dans les autres genres, notamment en littérature et qui complique la reconnaissance de l'auteur d'un papier scientifique.

Le tableau en annexe met également en valeur une autre caractéristique de la littérature scientifique signalée dès l'introduction : le degré élevé d'intertextualité tolérée, assumée, voire encouragée…

Une forte intertextualité assumée

Les différences dans la variable auteur suffisent donc à expliquer les distances relativement élevées entre des papiers issus des travaux d'un collectif à géométrie variable sur un même thème (même si les sujets et les angles d'attaque peuvent varier). Est-ce que ces différences suffisent pour expliquer les discordances apparaissant à partir du 21e couple de textes ?

A priori, les différences dans la composition des équipes de recherche ne peuvent expliquer les distances anormalement faibles constatées avec des papiers – signés par des auteurs différents – sur un même sujet (en gras sur le tableau 4).

Le premier couple présente même une distance légèrement inférieure au seuil le plus rigoureux ($\alpha = 0.3\%$), c'est-à-dire quasiment improbable du moins selon les hypothèses du modèle d'attribution d'auteur présenté plus haut. A la lecture, il apparaît que le dernier paru est un article de synthèse, portant exactement sur le même sujet avec la même terminologie, le même plan, des notations identiques, des conclusions très semblables.

Dans la littérature scientifique, ce type d'articles – de synthèse ou "survey paper" est fréquent. Les éditeurs les acceptent volontiers car ils ne posent pas de problèmes et leur évaluation est aisée, en tous cas plus aisée que pour un papier novateur... Les auteurs qui font l'objet de cette synthèse la tolèrent volontiers - même s'ils sont plus ou moins démarqués… à condition d'être cités. Est-il nécessaire de rappeler, à ce sujet, que la productivité d'un chercheur n'est pas tant mesurée à l'aune de ses publications que du nombre de fois où elles sont citées ?

Le corpus O fournit plusieurs exemples de cette intertextualité tolérée, voire encouragée, sur des sujets assez standardisés. Le numéro O0259[5] fournit le plus remarquable avec 10 couples de textes "trop" proches (tableau 5). Dans le tableau en annexe, il apparaît dans 6 couples pour des distances remarquablement faibles avec $\alpha = 0.3\%$).

---

[5] Marianne A.Azer, Sherif M. El-Kassas, Magdy S. El-Soudani. A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks. *International Journal of Computer Science and Information Sec*urity, Special Issue, May 2009.

Tableau 5 Couples remarquables formés avec l'article de synthèse O0259 paru en 2009 ($\alpha < 1\%$)

Rang du couple A,B	Texte B	Distance (A,B)	Année parution de B	Longueur de B (nombre mots)
22	O0260	0,4218	2006	11 916
29	O0024	0,4356	2003	6 627
32	O0110	0,4491	2005	4 460
33	O0233	0,4598	2000	4 886
40	O0102	0,4690	2004	3 121
50	O0264	0,4797	2003	6 927
76	O0144	0,4938	2004	5 148
95	O0115	0,4986	2005	3 334
120	O0078	0,5026	2004	7 740
127	O0169	0,5035	2005	3 654

Ce cas appelle les remarques suivantes :

- l'article de synthèse est long (7116 mots). De plus, le couple au rang 22 concerne un papier hors limite (longueur supérieure à la limite supérieure de l'intervalle défini ci-dessus). Ce facteur longueur a pu également jouer pour les couples de rangs 29, 50, 120…

- les années de parution indiquent clairement le sens des influences. Ce sont tous des articles faisant le tour de la même question à des dates successives. Ils fournissent à O0259 une bonne partie de la « matière première », notamment la terminologie et le symbolisme, sans compter les références qui, rappelons-le n'entrent pas dans le calcul.

Cet exemple permet de comprendre que ces textes anormalement proches sont unis par une forte intertextualité. Mais cette intertextualité ne semble pas franchir la limite de ce qui est implicitement toléré – voire encouragé - dans le domaine de la publication scientifique.

En conclusion sur ce point, dans le tableau en annexe, on trouve : 28 couples de textes ayant – au moins partiellement - le ou les mêmes auteurs (sur un total de 32), la distance moyenne "intra" est de .3650 ; et 25 couples de textes "anormalement" proches bien qu'ayant des auteurs différents. Pour ces derniers, la distance moyenne "inter" est de .4595, soit 26% plus élevée que la moyenne "intra". On retiendra surtout que seulement 0,075% des distances entre textes d'auteurs différents sont inférieures à 0.49. Autrement dit, en retenant ce seuil, on aurait moins de 1 chance sur mille de se tromper en affirmant que deux textes séparés par une distance inférieure à 0.49 sont bien du même auteur.

L'idée selon laquelle le facteur auteur est bien prépondérant ne peut donc être rejetée. Cependant, ce facteur se trouve en partie contrecarré par d'autres caractéristiques propres aux publications scientifiques. De ce fait, il n'est pas possible de raisonner de manière dichotomique. Ce n'est que dans le cas de distances très faibles que l'on peut conclure avec certitude que l'auteur est le même. Ensuite, il existe une zone grise où plusieurs hypothèses sont à considérer et à valider par examen des textes. L'une est très probable : auteurs en tout ou partie semblables. Une autre hypothèse rare mais possible : un article de synthèse (B) partageant, avec un texte (A), une proportion élevée de texte commun, sans que cette intertextualité pose problème étant donné les principes de la publication scientifique…

Tel n'est évidemment pas le cas lorsque cette intertextualité est dissimulée.

II. L'intertextualité dissimulée

Le corpus O est composé de documents originaux (d'où "O") qui ont un point commun : tous ont été signalés par l''IEEE comme ayant été "dupliqués en violation des principes de la publication scientifique" par d'autres auteurs (ces documents dupliqués sont identifiés par la lettre D. La liste complète est présentée en annexe 1.

Duplication sans référence ni permission.

Dans la base de l'IEEE, nous avons identifié – au printemps 2012 – plus de 300 textes précédés d'un avertissement semblable à celui reproduit dans le tableau 6 ci-dessous.

Tableau 6 Exemple d'avertissement placé en tête d'un article présent dans la base bibliographique de l'IEEE

Notice of Violation of IEEE Publication Principles
"Estimating neutral divergence amongst Mammals for Comparative Genomics with Mammalian Scope" by Anup Bhatkar and J.L. Rana in the Proceedings of the 9th International Conference on Information Technology (ICIT'06)
After careful and considered review of the content and authorship of this paper by a duly constituted expert committee, this paper has been found to be in violation of IEEE's Publication Principles.
This paper contains significant duplication of original text from the papers cited below.
The original text was copied without attribution (including appropriate references to the original author(s) and/or paper titles) and without permission.
Due to the nature of this violation, reasonable effort should be made to remove all past references to this paper, and future references should be made to the following articles:
"Distribution and intensity of constraint in mammalian genomic sequence" by Gregory M. Cooper, Eric A. Stone, George Asimenos, Eric D. Green, Serafim Batzoglou, and Arend Sidow in Genome Research, Jul 2005; 15, pp 901 – 913, Cold Spring Harbor Press.
and
"Quantitative Estimates of Sequence Divergence for Comparative Analyses of Mammalian Genomes" by Gregory M. Cooper, Michael Brudno, Eric D. Green, Serafim Batzoglou, and Arend Sidow in Genome Research, May 2003; 13, pp 813 – 820, Cold Spring Harbor Press

Les termes utilisés par l'IEEE donnent une définition assez précise du phénomène : la **copie d'un ou plusieurs textes, sans indiquer les références de ces originaux et sans la permission** du ou des auteurs desdits originaux. Nous proposons d'appeler ce phénomène : "intertextualité dissimulée".

Ce phénomène soulève deux questions :

Premièrement, à partir de quel moment peut-on parler de duplication ? Dans l'avertissement de l'IEEE l'adjectif "significatif" pose problème : à partir de quelle proportion de texte recopié pourra-t-on utiliser le terme de "duplication" ? En fait, cette appréciation a été laissée aux comités d'experts auxquels l'IEEE a soumis les cas signalés à son attention. Elle est d'autant plus délicate que, comme nous l'avons signalé, l'intertextualité est l'un des

phénomènes caractéristiques de la littérature scientifique… D'où cette question : les cas relevés par l'IEEE permettent-ils de déterminer une sorte de "jurisprudence" ?

Deuxièmement, l'informatique pourrait-elle aider à détecter ces "duplications sans référence ni permission", c'est-à-dire cette intertextualité dissimulée ? Cette idée n'est pas neuve – synthèse dans Bouville 2008 - et de nombreuses méthodes de détection ont déjà été proposées (Alzahrani et Al 2011). Celle que nous présentons aujourd'hui est, à notre connaissance, inédite.

Détection des proximités entre originaux et duplications

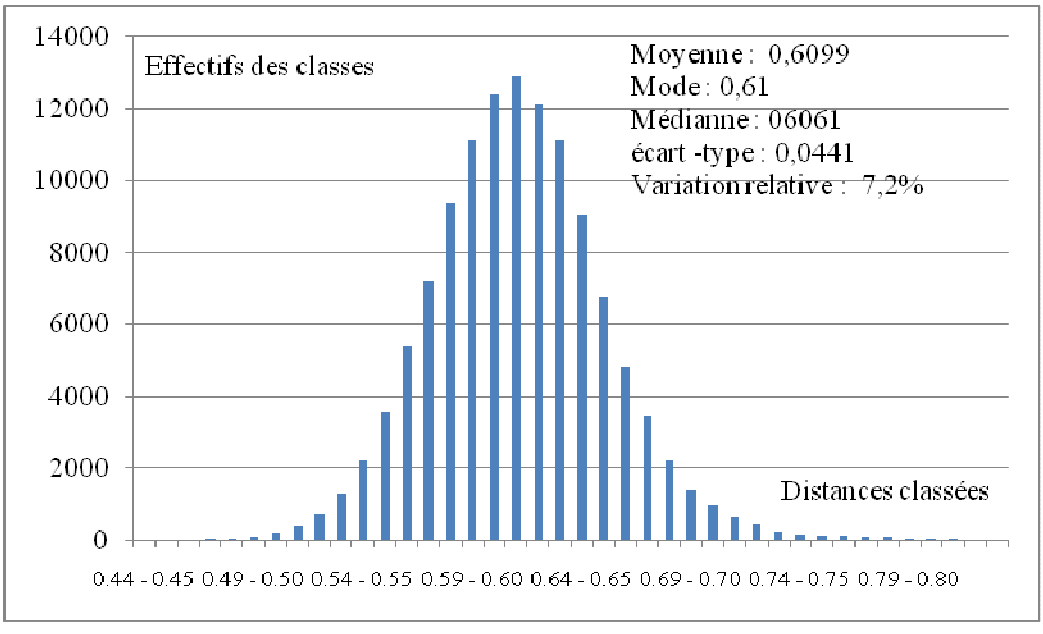
Pour répondre à ces deux questions, nous avons constitué un corpus D avec les textes précédés de l'avertissement dans la base de l'IEEE[6]. Ils sont plus de 300, mais 236 seulement ont pu être exploités. Pour certains les fichiers .pdf étaient des images impossibles à traiter en OCR ; pour d'autres, les originaux n'étaient pas disponibles ou illisibles…

Ces 236 duplications sont ajoutées aux 256 originaux sur lesquels elles auraient été, en tout ou partie, copiées (selon l'IEE) (voir Annexe 1).

Les mêmes procédures que ci-dessus sont appliquées à ces textes : conversion du fichier pdf en txt, segmentation en tokens, indexation puis calcul des distances entre les 478 textes dont la longueur est au moins égale à 1000 mots[7].

Le tableau 7 ci-dessous donne les résultats de cette opération et peut se comparer au tableau 2.

Tableau 7 Histogramme des distances dans le corpus O-D classées par ordre croissant (intervalle de classe : 0.01)



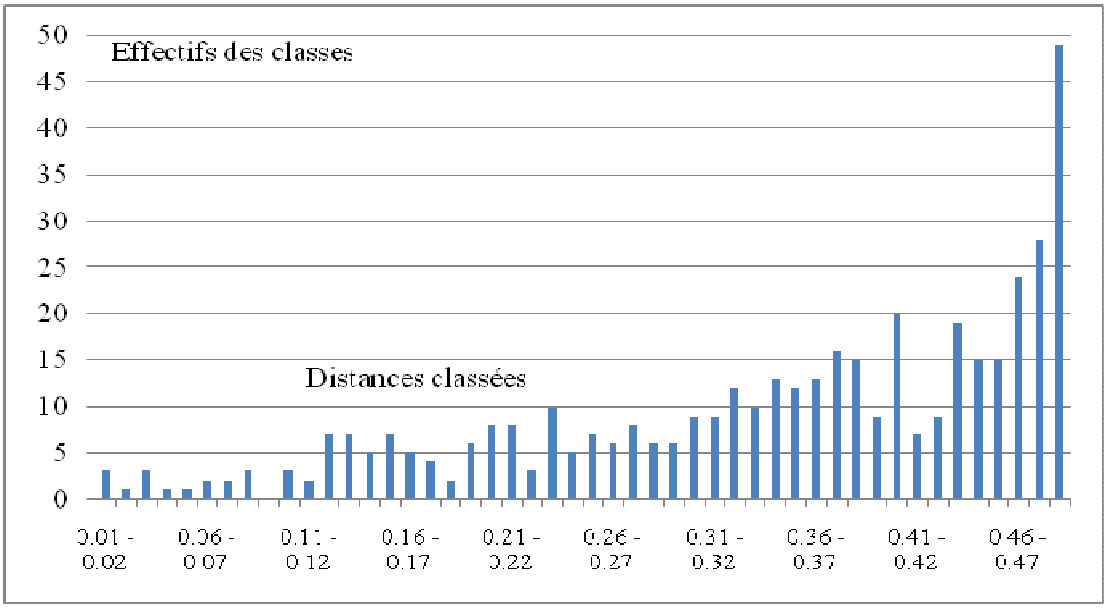
[6] Une première expérience limitée a été présentée dans Labbé & Labbé 2002b.

[7] Quatre dupliqués ont moins de 1000 mots : D0031, D0044, D0104 et D0057 : on ne mesure pas leur distances mais ils sont conservés pour la dernière partie de l'expérience (la fenêtre glissante).

On trouve la même courbe en cloche, des valeurs centrales semblables et confondues et à peu près les mêmes paramètres de dispersion. Cela suggère que le corpus total présente bien les caractéristiques d'un échantillon aléatoire : chaque décision de duplication est indépendante des autres et ce grand nombre de décisions indépendantes aboutit au même résultat qu'une sélection au hasard.

Mais ici un phénomène nouveau apparaît : un nombre trop élevé de distances anormalement faibles. En s'en tenant à la borne inférieure de l'intervalle de confiance avec $\alpha = 0{,}3\%$, elles sont trois fois plus nombreuses qu'attendues. Certaines de ces distances sont quasiment nulles et elles forment une "queue de distribution" tout à fait anormale (histogramme ci-dessous).

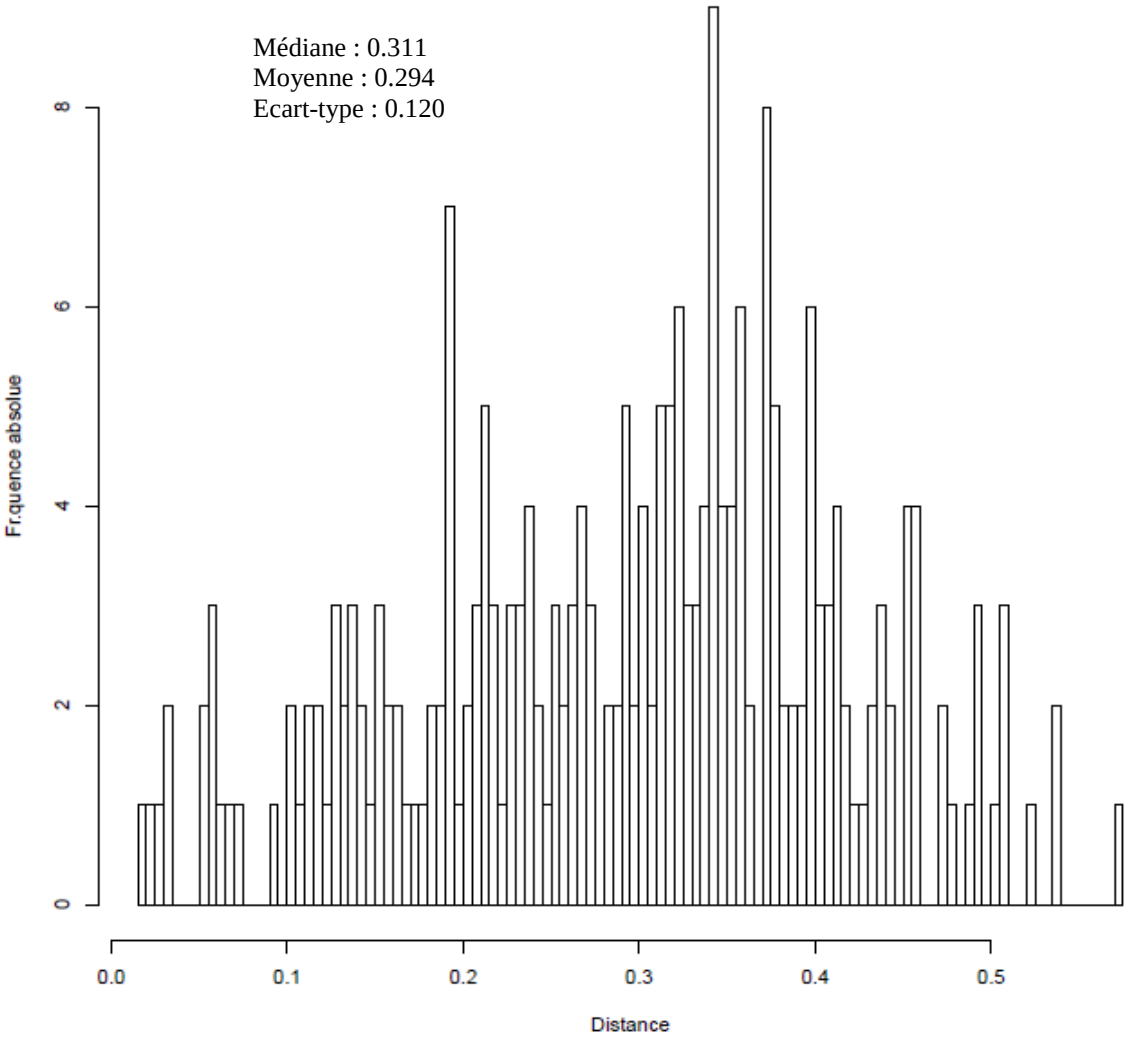
Tableau 8 Histogramme des distances remarquablement faibles ($\alpha = 0.3\%$, intervalle de classe : 0.01)



Dans ces distances remarquablement faibles, on trouve la plupart des duplications signalées par l'IEEE.

On peut le contrôler, en confrontant duplications et originaux et en ne retenant que les distances minimales entre les uns et les autres (tableau 9).

Tableau 9. Distances entre chaque duplicat et son original le plus proche (intervalle de classe : 0.01)



Par rapport aux graphiques (2 et 7), les valeurs centrales très basses indiquent des distances beaucoup trop faibles pour des textes supposés être d'auteurs différents. La forte dispersion des distances autour de la moyenne – comme le profil multimodal du graphique - signalent également une population fortement hétérogène.

Pour automatiser la détection de ces cas anormaux, plusieurs techniques de classification peuvent être utilisées. La plus simple consiste à rechercher le ou les plus proches voisins : "nearest neighbors classification" (knn classification, Cover and Hart 1967), en fixant un seuil d'acceptation, par exemple toute distance inférieure à 0,49 qui revient à accepter un risque d'erreur inférieur à 3‰.

Il apparaît nécessaire de chercher au-delà du premier voisin (k > 1) – même si cela complique sérieusement les choses - car, dans le corpus D, un nombre important de duplications ne portent pas sur un seul texte mais sur plusieurs et il est important de retrouver le maximum de ces duplications partielles, en ne s'arrêtant pas seulement à la plus flagrante.

L'annexe 3 reproduit les résultats de cette classification avec k = 3. Les dimensions du tableau ne permettent pas d'aller au-delà, mais on remarquera que ce n'est pas nécessaire, à ce stade, car très peu de duplications concernent plus de 3 textes originaux.

En retenant le seuil $\alpha = 0{,}3\%$ - 3 chances sur 1000 de se tromper en affirmant que les deux textes ont les mêmes auteurs ou que l'un des deux duplique l'autre d'une manière anormale – seules 6 duplications (sur 236) échappent à l'algorithme, soit 2,5% des cas recensés par l'IEEE. En fait, la réalité est nettement inférieure à la limite théorique énoncée ci-dessus puisqu'il n'y a que 0,7 ‰ faux positifs (moins de 1 erreur pour mille cas examinés). A ce propos, on remarquera que le but recherché n'est pas tant de retrouver toutes les duplications mais d'abord d'éviter au maximum les erreurs de diagnostic ("faux positif") en acceptant pour cela un risque assez faible de laisser échapper quelques cas.

Naturellement, on ne compte pas comme "faux positifs" les "effets de chaîne" entraînés par les proximités "anormales" entre originaux, notamment les autres textes sur les mêmes sujets par les mêmes auteurs ou les survey papers détectés dans la première section de cette communication. Par exemple,

- selon l'IEEE, D0014 a dupliqué les originaux 00018 et O0024. Effectivement, O0018 apparaît bien comme le plus proche voisin de D0014 avec une distance telle que la duplication ne fait pas de doute. Mais comme l'indique le tableau 4 ci-dessus, les auteurs de O0018 (D. J. McArthur et autres) ont quatre autres textes dans le corpus des originaux dont O0034 et O0035 sont extrêmement proches de O0018. Il est donc logique que O0034 et O0035 apparaissent en deuxième et troisième voisins de D0014 – avant même O0024 – sans qu'il s'agisse pour autant de « faux positifs »…
- D0008 a dupliqué O0024 et l'IEEE ne donne que ce texte original comme source. Cependant, ce texte étant l'un de ceux dont rend compte le survey paper O0259, il est logique que celui-ci soit détecté comme deuxième voisin de D0008.

Comme l'indique le profil du graphique ci-dessus et l'écart type élevé, il ne s'agit pas d'une population homogène. Trois catégories principales peuvent être distinguées.

Trois catégories principales de duplications

Premièrement, *la copie intégrale* : le texte original a été dupliqué sans aucune modification ou avec quelques changements cosmétiques, notamment dans le titre et le résumé.

Par exemple, le texte : D0173 Tang Shu-Pian. Resonant Energy Transfer from Organics to Quantum Dots and Carrier Multiplication. *Proceedings of the 2009 Asia-Pacific Power and Energy Engineering Conference*, March 2009

est une copie intégrale – sans aucune modification - de :

O0197 : Vladimir M. Agranovich & Gerard Czajkowski. *Resonant Energy Transfer from Organics to Quantum Dots and Carrier Multiplication*. (texte déposé sur arXiv en janvier 2008).

La distance de 0,018 s'explique par quelques fautes dans cette copie et par des différences de mise en page… Il peut aussi se produire quelques erreurs d'OCR.

Certains n'hésitent pas à répéter l'opération plusieurs fois. Par exemple, D0203 et D0219 sont un seul texte[8] – intégralement copié sur O0232[9] – présentés, sous le même titre, à deux conférences différentes[10].

[8] Jianfeng Zheng & Zaiqing Nie. Language Models for Web Object Retrieval.

Deuxième catégorie : *la copie de passages entiers* sans modifications notables avec quelques raccourcis et chevilles entre les passages originaux, ce qui donne un texte dupliqué plus bref et en "tissu d'Arlequin".

Par exemple : D0203 (Jianfeng Zheng & Zaiqing Nie. Language Models for Web Object Retrieval. *Proceedings of the International Conference on New Trends in Information and Service Science*, June 2009, pp. 282-287.

Ce "duplicat" compte 2700 mots. Il provient de :

O0232 (Zaiqing Nie, Yunxiao Ma, Shuming Shi, Ji-Rong Wen & Wei-Ying Ma. Web Object Retrieval. *Microsoft Research Technical Report*: MSR-TR-2006-70).

Qui est long de 4 906 mots. Les raccourcis et "coutures" entre passages copiés engendrent une distance de 0,240.

Dans ce cas, les mêmes ont présenté exactement le même texte (D0219) devant la *5th International Conference on Wireless Communications, Networking and Mobile Computing* (September 2009) !

Cette "reduplication" semble assez courante. Dans le corpus on trouve même un groupe d'auteurs qui ont présenté 4 fois quasiment le même texte dupliqué de cette manière[11].

Il se produit même parfois d'étranges collisions. Par exemple, l'article O0190[12], publié en 2007, est dupliqué en "tissu d'Arlequin" par deux auteurs qui le présentent en février 2009[13]. Cette première duplication est, à son tour, presqu'intégralement dupliquée par un autre groupe qui la présente, exactement sous le même titre, en juin 2009[14]…

Il arrive assez souvent que la duplication ne porte pas sur un texte mais sur deux – voir l'avertissement reproduit ci-dessus - voire plus : l'un des cas dénoncé par l'IEEE est un collage de 5 textes différents. Ces cas plus complexes génèrent parfois des distances assez importantes entre certains originaux et leur duplication. C'est également le cas pour la troisième catégorie.

Troisième catégorie : *l'insertion de passages plus ou moins remaniés* dans une paraphrase, voire une création plus ou moins personnelle. Ce sont les cas les plus difficiles à

---

[9] Zaiqing Nie, Yunxiao Ma, Shuming Shi, Ji-Rong Wen & Wei-Ying Ma. *Web Object Retrieval. Microsoft Research Technical Report*: MSR-TR-2006-70

[10] *International Conference on New Trends in Information and Service Science (*June 2009) et *5th International Conference on Wireless Communications, Networking and Mobile Computing (*September 2009).

[11] L. Gatani, G. Lo Re, S. Gaglio. D0062 : An Adaptive Routing Protocol for Ad Hoc Peer-to-peer Networks. *Proceedings of the Sixth International Symposium on World of Wireless Mobile and Multimedia Network*s, 2005. pp. 44-50, 13-16 June 2005. D0067. Efficient Query Routing in Peer-to-peer Networks. *Proceedings of the Third International Conference on Information Technology: Research and Education*, 2005. ITRE 2005. pp. 393-397, 27-30 June 2005. D0069. Reinforcement Learning for P2P Searching. *Proceedings of the Seventh International Workshop on Computer Architecture for Machine Perception*, 2005. pp. 303-308, 4-6 July 2005. D0071 (01558556) L. Gatani, G. Lo Re, S. Gaglio. An Adaptive Routing Mechanism for P2P Resource Discovery. *Proceedings of the IEEE International Symposium on Cluster Computing and the Grid*, 2005. pp. 205-212 Vol. 1, 9-12 May 2005.

[12] S. Karetsos, C. Costopoulos, O. Pyrovolakis & L. Georgiou. Developing Agricultural B2B Processes Using Web Services. *Proceedings of the 6th WSEAS Int. Conf. on Software Engineering, Parallel and Distributed Systems*, February 16-19, 2007, pp. 161-168.

[13] D0167 Chenxia Sun, Xiaoyang He, Chao Wang. The Web-based B2B Environment with Web Services. *Proceedings of the 2009 International Conference on Electronic Computer Technology*, February 2009, pp. 103-107.

[14] D0175 Xinyu Zhang, Wei Tang, Cheng Sun. The Web-based B2B Environment with Web Services. *Proceedings of the 2009 Sixth International Conference on Information Technology: New Generations*, June 2009, pp. 1623-1624.

détecter car les distances sont généralement plus élevées. Dans trois cas, elles excèdent même le seuil fixé avec α = 1%. Mais, dans deux de ces cas, il est possible que la version de l'original disponible en ligne soit assez différente de celui qui aurait été utilisée pour la duplication.

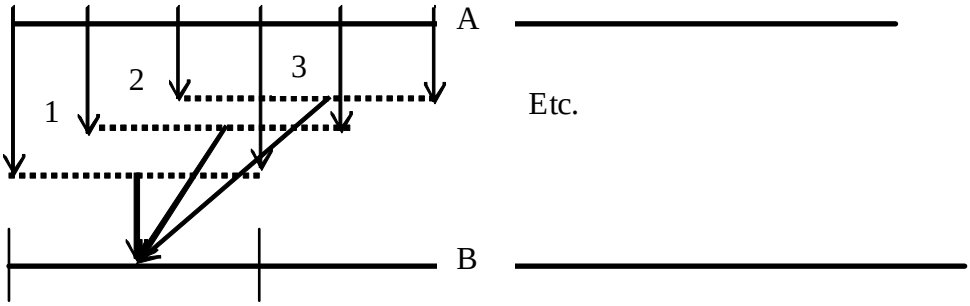
Cela illustre les difficultés inhérentes à la constitution de vastes corpus de qualité. Cela montre aussi combien il est difficile de tirer une "jurisprudence" claire et simple des cas repérés dans la base de l'IEEE.

Pour cette troisième catégorie, il se pose une question évidente : comment détecter les passages dupliqués ?

Détection des passages dupliqués : la fenêtre glissante

Lorsqu'une proximité anormale est détectée par la méthode présentée ci-dessus, il est proposé de procéder de la manière suivante. Chacun des deux textes est divisé en petites fenêtres de longueurs égales qui sont comparées deux à deux (tableau 10). Cette méthode a été discutée dans (Labbé 2007). Une méthode similaire est également présentée dans (Brixtel & Al. 2009).

Tableau 10 La méthode de la fenêtre glissante (un fragment de A est comparé à tous les fragments de même longueur découpés selon un même pas dans B).



Au départ, la méthode est la même que celle présentée dans la première section : calcul d'une moyenne, d'un intervalle de confiance et détection des proximités anormales – mais elle est appliquée à tous les fragments de même dimension que l'on peut découper dans les deux textes.

Voici d'abord un exemple avec une fenêtre de 250 mots se déplaçant le long du texte selon un pas de 125 mots (autrement dit chaque mot des deux textes intervient deux fois dans les divers calculs. La distance moyenne entre toutes les fenêtres de 250 mots découpées entre textes d'auteurs différents dans le corpus O : 0.687 ; écart type : 0.040 ; limite inférieure de l'intervalle de confiance (α = 1%) : 0,562.

On peut considérer qu'une distance inférieure à ce seuil de similarité (0,56) entre deux fenêtres de 250 mots découpées dans des textes signés par des auteurs différents doit être examinée avec attention. Le tableau 11 ci-dessous donne un résumé de ces résultats obtenus sur les trois textes cités dans l'avertissement reproduit au début de cette section (pour compresser le tableau, les fenêtres contigües ont été fusionnées quand elles étaient significatives).

Tableau 11 Détection – dans un texte unique - des portions copiées sur plusieurs originaux différents grâce à la fenêtre glissante

Fenêtre(s)	Dupliqué (D0002)	Originaux	Portions des originaux	Plus petite distance
1	0 - 250	O0011	0 - 750	0.38
2	250 - 500	O0011	500 - 1250	0.42
3	500 - 750	O0011	1000 - 1250	0.52
4	750 - 1000	O0003	6750 - 7250	0.50
5	1000 - 1500	O0003	7250 - 7750	0.25
6	1500 - 1600	O0011	2000 - 2250	0.54

La tranche numéro 5 est la plus proche de l'un des deux originaux dupliqués. Les deux textes peuvent être lus en parallèle dans l'annexe 4. Naturellement, dans ce cas précis, la duplication est grossière mais il est intéressant de noter que la combinaison de la fenêtre glissante avec la distance intertextuelle et la méthode des seuils permet de localiser automatiquement les zones concernées et de mettre en parallèle les extraits de l'original et de la duplication, ce qui peut faciliter considérablement la tâche des experts à qui seraient soumis ces cas.

Validation et détermination des paramètres optimaux

Comme ci-dessus, deux questions sont posées. Premièrement, comment valider la méthode ? Deuxièmement, quelles sont les dimensions optimales pour obtenir le plus faible risque possible de faux positif, tout en assurant un taux de reconnaissance le plus élevé possible sans emcombrer l'opérateur avec une trop grande masse de texte ?

Pour valider la méthode et déterminer ces dimensions optimales, nous avons fait varier la taille de la fenêtre, du pas et du seuil de similarité, en décomptant à chaque fois l'ensemble des résultats. Comme précédemment, les textes retenus ont une longueur supérieure à 1000 mots. Pour qu'un couple soit étudié, il faut que le rapport des deux longueurs soit inférieur à 1 :7 et qu'ils soient séparés par des distances inférieures à 0,49.

Il apparaît d'abord que plus la fenêtre est petite – à seuil de similarité égal - plus le risque de "faux positif" augmente : certaines des tranches suspectes ne sont pas des duplications mais, par exemple, des portions de textes portant sur des aspects techniques identiques ou sur les mêmes publications antérieures, etc. Avec une fenêtre inférieure à 250 mots, ce risque devient non négligeable. La longueur qui génère le moins d'erreur est celle de 500 mots.

Au passage, cette constatation rappelle qu'il semble impossible – par des méthodes statistiques - de retrouver l'auteur d'une lettre de quelques lignes ou d'un poème de quelques vers. C'est aussi pour cela que la technique de la fenêtre glissante ne peut servir à la détection de la duplication mais seulement à la localisation des passages concernés quand cette duplication a déjà été détectée sur l'ensemble des textes.

Le pas a une incidence en cas de duplication en "tissu d'Arlequin" avec des extraits de longueur limitée. Dans ce cas, un pas trop important risque de laisser passer certains passages si aucune fenêtre ne les contient sans trop de chevauchements avec des parties non dupliquées. En contrepartie, diminuer le pas ralentit l'exécution et rend plus difficile le dépouillement des résultats…

Le seuil de similarité est le dernier paramètre important. Le tableau 12 permet de mesurer son impact. Ce sont les résultats obtenus sur les 267 couples duplication/original séparés par une distance intertextuelle inférieure à 0,49. Les deux premières colonnes donnent les résultats avec un seuil très sévère (0.2 ou 80% du texte en commun). Les deux suivantes avec le seuil le plus élevé (0,50).

Tableau 12. Proportion (P) des tranches semblables détectées dans un couple original/ duplication en fonction du seuil de similarité (fenêtre de 500 mots et pas de 50 mots)

	Seuil 0.2		Seuil 0.5	
Proportion de tranches semblables	Nombre de couples	Distance moyenne	Nombre de couples	Distance moyenne
Toutes	11	0,0794	125	0 ,2383
1 > P > 0,75	13	0.0907	38	0,3190
0.75 > P > 0,50	28	0,1864	48	0,3552
0,50 > P > 0,25	29	0,2421	35	0,3937
0,25 > P > 0	69	0.3109	19	0,4327
0	58	0,3841	1	0,4425

En définitive, le choix du seuil et du pas dépendent de la catégorie à laquelle appartient la duplication et de la stratégie choisie pour sa détection :

- un seuil bas permet d'aller directement aux passages intégralement ou quasi-intégralement dupliqués. Cette stratégie s'impose en cas de distance intertextuelle faible entre un original et une duplication potentielle ou lorsque plusieurs originaux sont détectés. En contrepartie, ce seuil est manifestement inadapté lorsque la distance est assez élevée comme en témoigne les 58 cas pour lesquels aucun résultat n'est retourné, alors qu'il s'agit bien de duplications.

- à l'opposé, un seuil de similarité plus élevé permet de retrouver les passages contenant une dose significative de duplication, passages qui peuvent être dispersés dans un texte qui n'est que partiellement une copie de l'original. En contrepartie, ce choix est contreproductif lorsque la distance est relativement faible : dans ce cas, la totalité (ou la majorité) du texte est signalé comme duplication…

Cette technique permet ainsi d'isoler des exemples significatifs - comme celui présenté dans l'annexe 4 - et de définir, pour certains paramètres, des proportions limites. Cela serait un premier pas vers la constitution d'une "jurisprudence" de la duplication. Toutefois, cette jurisprudence sera toujours incomplète puisque l'IEEE ne publie pas les demandes qu'elle rejette et qui seraient indispensables pour définir le "taux d'intertextualité" acceptable, en face de l'inacceptable qui nous est donné à voir…

Conclusions

Par rapport à l'auteur littéraire, l'auteur scientifique peut se révéler plus difficile à identifier. Plusieurs facteurs expliqueraient cette difficulté occasionnelle, comme l'impersonnalité du propos ou la géométrie variable des auteurs collectifs et surtout une proportion importante d'intertextualité acceptée, voire recherchée, entre papiers portant sur un même sujet. On pourrait d'ailleurs se demander si cela n'est pas une caractéristique même du développement de la science. Quand un champ de recherche se dirige vers la maturité, terminologie, symbolisme, expériences et exemples se standardisent progressivement, de telle sorte que les papiers sur ces sujets partageraient une dose d'intertextualité de plus en plus forte. L'informatique, science jeune, donne l'occasion d'observer ce processus, notamment avec les nombreux "survey papers" qui contribuent grandement à cette convergence. Tout cela augmenterait le taux d'intertextualité acceptée et rendrait plus difficile l'identification de l'auteur, du moins dans quelques cas...

Sous ces réserves, les expériences présentées dans cette communication suggèrent que la combinaison de la distance intertextuelle, de la fenêtre glissante et de diverses techniques de classification fournit un outil adapté pour détecter les cas d'intertextualité dissimulée dans la littérature scientifique.

L'enjeu est d'importance. Etablir l'origine des idées, des concepts, des algorithmes et des données n'est pas une question morale mais une condition du progrès scientifique. C'est à cette condition que les chercheurs acceptent de communiquer les résultats de leurs travaux. L'intertextualité cachée, en suscitant la méfiance, freine les échanges et complique la recherche (par exemple : Arnold 2010).

Des logiciels performants, capables de détecter cette intertextualité dissimulée, seraient donc très utiles. L'une des difficultés rencontrées dans la mise au point de ces logiciels de détection est l'absence d'un corpus étendu de cas réels sur lesquels tester ces outils. La base bibliographique de l'IEEE répond donc à ce besoin en fournissant plusieurs centaines de cas exploitables. Naturellement, il ne s'agit pas de stigmatiser des individus. Telle est la raison pour laquelle il faut s'en tenir à "duplication" et ne pas utiliser des termes comme "plagiat", "plagiaire" ou "fraude".

Dans le futur, ces outils de "data mining" pourraient certainement être utiles aux organisateurs de conférences, aux éditeurs de revues, aux gestionnaires des bases bibliographiques, aux commissions de spécialistes pour contrer certaines pratiques. Naturellement ces automates ne seront que des aides à la décision ; celle-ci dépendra toujours *in fine* d'une lecture attentive – par des experts du domaine - des cas repérés par la machine. Cette précaution est d'autant plus nécessaire que – comme cette communication l'a montré - le profil de l'auteur collectif est souvent à géométrie variable mais aussi parce que certaines pratiques peuvent générer une intertextualité importante quoique considérée comme normale, du moins en l'état actuel des choses.

Naturellement, certaines formes d'intertextualité dissimulée resteront hors d'atteinte. C'est en particulier le cas lorsque le document dupliqué est dans une autre langue que la duplication ou que l'"emprunt" dissimulé porte non pas sur le texte mais sur les données, le symbolisme, les figures…

Enfin, soulignons que ces expériences montrent que l'auteur d'un texte scientifique est identifiable, tout comme l'est celui d'un texte littéraire, journalistique ou politique.

Bibliographie

Alzahrani S., Salim N. & Abraham A. (2011). "Understanding Plagiarism Linguistic Patterns, Textual Features and Detection Methods". *IEE Transactions on Systems, Man and Cybernetics – Part C: Applications and Reviews*, (PP 99).

Arnold D. N. (2010). Integrity Under Attack:The State of Scholarly Publishing. *SIAM News. Journal of the Society for Industrial and Applied Mathematics*. 42-10, Décembre 2010 (consultable en ligne sur le site de SIAM).

Bouville M. (2008). Plagiarism: Words and ideas. *Science and Engineering Ethics*, 14(3), 311-322.

Brixtel R., Lesner B., Bagan G. & Bazin C. (2009). De la mesure de similarité de codes sources vers la detection de plagiat : le “Pomp-O-Mètre”. MajecSTIC 2009.

Cover T. M. & Hart P. E. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*. 13, 21–27.

Hockey S. & Martin J. (1988). *OCP Users’ Manual*. Oxford: Oxford University Computing Service.

Koppel Moshe, Schler Jonathan & Argamon Shlomo (2009). Computational Methods in Authorship Attribution. *Journal of the American Society for Information Science and Technology*. 60-1, 9-26.

Labbé Cyril & Labbé Dominique (2003). La distance intertextuelle. *Corpus*, 3, p. 95-118.

Labbé Cyril & Labbé Dominique (2006). A Tool for Literary Studies: Intertextual Distance and Tree Classification. *Literary and Linguistic Computing*. 21-3, 2006, p. 311-326.

Labbé Cyril & Labbé Dominique (2011). La classification des textes. Comment trouver le meilleur classement possible au sein d’une collection de textes ? *Images des mathématiques. La recherche mathématique en mots et en images*. (http://images.math.cnrs.fr/La-classification-des-textes.html). 28 mars 2011.

Labbé Cyril & Labbé Dominique (2012a). Duplicate and fake publications in the scientific literature: how many SCIgen papers in computer science? *Scientometrics*. 22 june 2012.

Labbé Cyril & Labbé Dominique (2012b). Detection of Hidden Intertextuality in the Scientific Publications. In Dister Anne, Longrée Dominique, Purnelle Gérald (éds). *Proceedings of the 11th International Conference on Textual Data Statistical Analysis*. Liège : LASLA - SESLA, p.537-551.

Labbé Dominique (2007). Experiments on Authorship Attribution by Intertextual Distance in English. *Journal of Quantitative Linguistics*. 14(1), 33–80.

Love Harold (2002). *Attributing Authorship: An Introduction*. Cambridge : Cambridge University Press.

Savoy Jacques (2012). Authorship Attribution: A Comparative Study of Three Text Corpora and Three Languages. *Journal of Quantitative Linguistics*. 19(2): 132-161.

Stamatatos Efstathios (2009). A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*. 60-3, p. 538-556.

Annexe 1. Le corpus des originaux (O) et des duplications (D) signalées par l'IEEE et qui ont pu être utilisés dans notre communication (liste établie au printemps 2012)

	Duplications		Originaux
D0001	T. J. Hammons. Status of International Interconnections and Electricity Deregulation in Africa. *Proceedings of the 10th International Workshop on Web and Databases*, June 15 2007, Beijing, China.	O0004	P Naidoo, L. Musaba, W Balet & A Chikova. Toward Developing a Competitive Market for Regional Electricity Cross Border Trading : the Case of the Southern African Power Pool. *IEEE 2004 General Meeting*, Denver, USA, June 2004
		O0005	A. Majeed, H A Karim, N.H Al Maskati, S. Sud. Status of Gulf Co-Operation Council (GCC) Electricity Grid System Interconnection. *Proceedings of the 2004 IEEE Power Engineering Society General Meeting*,
		O0006	Ahmed Zobaa. Status of International Interconnections. *Proceedings of the 2004 IEEE Power Engineering Society General Meeting*,
		O0007	Raymond Johnson. Impact of Privatization and Deregulation on Infrastructure Development in Africa. *Proceedings of the 2004 IEEE Power Engineering Society General Meeting.*
D0002	Anup Bhatkar & J.L. Rana, Estimating neutral divergence amongst Mammals for Comparative Genomics with Mammalian Scope. *Proceedings of the 9th International Conference on Information Technology*.	O0003	Gregory M. Cooper, Eric A. Stone, George Asimenos, NISC Comparative Sequencing Program, Eric D. Green, Serafim Batzoglou and Arend Sidow, Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, Jul 2005; 15, pp 901 – 913
		O0011	Gregory M. Cooper, Michael Brudno, NISC Comparative Sequencing Program, Eric D. Green, Serafim Batzoglou, and Arend Sidow, "Quantitative Estimates of Sequence Divergence for Comparative Analyses of Mammalian Genomes. *Genome Research*, May 2003; 13, pp 813 – 820.
D0003	Krzysztof Szafranski. Analysis of Hemodynamics of Intercranial Saccular Aneurysms. *Proceedings of the 29th Annual International Conference of the IEEE EMBS, 23,August 2007, pp 2859-2862*.	O0012	Yiemeng Hoi, Hui Meng, Scott H. Woodward, Bernard R. Bendok, Ricardo A. Hanel, Lee R. Guterman, and L. Nelson Hopkins.Effects of Arterial Geometry on Aneurysm Growth: Three-dimensional Computational Fluid Dynamics Study. *Journal of Neurosurgery*, October 2004, pp 676-681.
D0004	David I. Eromon. High Temperature Superconducting (HTS) Generator Field Coil with Influence of Thermal AC Losses.	O0002	NMagnusson and M Runde.The influence of thermal gradients on AC losses in high-temperature superconducting coils. *Supercond.*

	Proceedings of the 33rd Annual Conference of the IEEE Industrial Electronics Society, Nov. 2007, pp. 1280-1286		*Sci. Technol*. 15 (2002) 1113–1118
D0005	Rahul Choudhari, Ajay Choudhari, R. D. Choudhari. Increasing Search Engine Efficiency using Cooperative Web. *Proceedings of the 2008 International Conference on Computer Science and Software Engineering*. Wuhan, China, December 12, 2008	O0001	Jie Xu Qinglan Li Huiming Qu Alexandros Labrinidis. Towards a Content-Provider-Friendly Web Page Crawler. *Proceedings of the 10th International Workshop on Web and Databases*, June 15, 2007, Beijing, China
D0006	Hong Fei, Liu Rui, Bai Yu. Performance Evaluation of the Burstiness Impact with a Realistic IP Structure Model. *Proceedings of the 2nd IEEE International Conference on Computer Science and Information Technology*, Beijing, China, August 2009.	O0009	Chloé Rolland, Julien Ridoux, Bruno Baynat, Vincent Borrel. Using LiTGen, a realistic IP traffic model, to evaluate the impact of burstiness on performance. *SIMUTools 2008*, March 03-07, 2008, Marseille, France.
D0007	Umesh Sehgal, Kuljeet Kaur, Pawan Kumar. The Anatomy of a Large-Scale Hyper Textual Web Search Engine. *Proceedings of the Second International Conference on Computer and Electrical Engineering*, December 2009, pp. 491-495.	O0010	Sergey Brin, Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30 (1998), p 107- 117.
D0008	Baolin Sun, Hua Chen. An Intrusion Detection System for AODV. *Proceedings of the 10th International Conference of Engineering of Complex Computer Systems*, 2005.	O0024	Yang Tseng, Poornima Balasubramanyam, Calvin Ko, Rattapon Limprasittiporn, Jeff Rowe & Karl Levitt. A Specification-based Intrusion Detection System for AODVC. *Proceedings of the 2003 ACM Workshop on Security of Ad Hoc andSensor Networks*, Fairfax, VA USA, 31October 2003.
D0009	HuaiKou Miao and JunFeng Wu. Applying Formal Methods to Compositionality Description of Web Service. *Second International Conference on Internet and Web Applications and Services ICIW 2007*.	O0016 O0014	M. Solanki, A. Cau & H. Zedan. Introducing Compositionality in Web Service Descriptions. *Proceedings of the 10th IEEE International Workshop on Future Trends of Distributed Computing Systems*, 2004, pp 14-20.. M. Solanki, A. Cau & H. Zedan. Augmenting Semantic Web Service Description with Compositional Specifications. *Proceedings of the International World Wide Web Conference*, May 2004, pp. 544 - 552.
D0010	M. Aruna, M.P. Suguna Devi & M. Deepa. Measuring the Quality of Software Modularization using Coupling-Based Structural Metrics for an OOS System. *First International Conference on Emerging Trends in Engineering and Technology*.	O0020 O0019	Santonu Sarkar, Girish Maskeri Rama & Avinash C. Kak. API-Based and Information-Theoretic Metrics for Measuring the Quality of Software Modularization. *IEEE Transactions on Software Engineering*, Vol. 33, N°. 1, January 2007. Santonu Sarkar, Avinash C. Kak & N. S. Nagaraja. Metrics for Analyzing Module Interactions in Large Software Systems. *Proceedings of the 12th Asia-Pacific Software Engineering Conference*.
D0011	Dong Lingxun, Dou Lihua & Feng Heping. Hybrid Time-Optimal	O0017	Mario Vasak, Mato Baoti´c, Ivan Petrovi´c & Nedjeljko Peri´c.

	Predictive Control for Mechanical Systems with Backlash Nonlinearity. *IEEEASME International Conference on Mechatronics and Embedded Systems and Applications*, October 2008.		Hybrid Theory Based Time-Optimal Control of an Electronic Throttle. *IEEE-ISIE* 2005, June 2005, pp. 20-23.
		O0013	Mattias Nordin & Per-Olof Gutmanin. Controlling mechanical systems with backlash — a survey. *Automatica.38* (2002), pp. 1633–1649.
		O0021	P. Rostalski, T. Besselmann, M. Bari, F. Van Belzen & M. Morari. A hybrid approach to modelling, control and state estimation of mechanical systems with backlash. *International Journal of Control*, vol. 80, no. 11, pp. 1729-1740
D0012	K. Inderjeet, T. Kamal, M. Kulkarni, G. Daya & A. Prabhjyot. Adaptive OFDM Vs Single Carrier Modulation with Frequency Domain Equalization. *Proceedings of the International Conference on Computer Engineering and Technology*. ICCET	O0015	Andreas Czylwik. Comparison between Adaptive OFDM and Single Carrier Modulation with Frequency Domain Equalization. *Proceedings of the IEEE 47th Vehicular Technology Conference*, 1997, vol.2, pp. 865-869.
D0013	H.M. Khodr, Zita A. Vale & Carlos Ramos. Optimal Cost-Benefit for the Location of Capacitors in Radial Distribution Systems. *IEEE Transactions on Power Delivery*, Vol 24, No 2, April 2009.	O0022	H.M. Khodr, F.G. Olsinab, P.M. De Oliveira-De Jesus & J.M. Yustad. Maximum savings approach for location and sizing of capacitors in distribution systems. *Electric Power Systems Research*, 78 (2008), Elsevier, pp. 1192-1203.
D0014	Dejia Shi, Li Wang & Jing He. The Design of Multi-agent System in IDAPS Microgrid. *Proceedings of the International Asia Symposium on Intelligent Interaction and Affective Computing*, 2009, pp. 63-66	O0018	Victoria M. Catterson, Euan M. Davidson & Stephen D. J. McArthur. Issues in Integrating Existing Multi-agent Systems for Power Engineering Applications. *Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems*, Nov. 2005.
		O0023	M. Pipattanasomporn, H. Feroze & S. Rahman. Multi-agent Systems in a Distributed Smart Grid: Design and Implementation. *Proceedings of the IEEE/PES Power Systems Conference and Exposition*, 2009, pp.1-8.
D0015	Huang Yi. Zerotree Wavelet Coding Using Fractal Prediction *Proceedings of the 1998 International Conference on Communication Technology (ICCT'98)*, vol 1. 2491.	O0025	Khai Uwe Barthel, Sven Brandau, Wolfgang Hermesmeier & Guido Heising. Zerotree Wavelet Coding Using Fractal Prediction. *Proceedings of the 1st International Conference on Image Processing*, 1997, vol 2, pp 314-317.
D0016	V. Nath, Ruchika Kumari, B. N. Das & R. N. Gupta. Design of Ultra-low Power CMOS Cells for Temperature Sensors in VLSI. *Proceedings of the 18th Annual Symposium on Integrated Circuits and System Design*, ACM Press, 2005, pp. 202–206	O0026	Conrado Rossi & Pablo Aguirre. Ultra-low Power CMOS Cells for Temperature Sensors. *Proceedings of the 18th Annual Symposium on Integrated Circuits and System Design*, ACM Press, 2005, pp. 202–206.

D0017	Wei Dai & Peng Hu. Elastic Consistency of Distributed Virtual Environment. *International Conference on Information Science and Engineering (ICISE)*, Dec. 2009, pp. 3310-3314.	O0027	Hermann Schloss, Jean Botev, Alex H¨ohfeld, Ingo Scholtes, Peter Sturm & Markus Esch. Elastic Consistency in Decentralized Distributed Virtual Environments. *Proceedings of the International Conference on Automated Solutions for Cross Media Content and Multi-channel Distribution*, Nov. 2008. pp.249-252.
D0018	Lei Li & Jianhao Hu. Redundant Residue Number Systems Based Radiation Hardening for Datapath. *IEEE Transactions on Nuclear Science*, vol.57, no.4, Aug. 2010, pp.2332-2343	O0049	Vik Tor Goh & Mohammad Umar Siddiqi. Multiple Error Detection and Correction Based on Redundant Residue Number Systems. *IEEE Transactions on Communications*, Vol.. 56, N°. 3, March 2008.
		O0050	Hari Krishna, Kuo-Yu Lin & Jenn-Dong Sun. A Coding Theory Approach to Error Control in Redundant Residue Number Systems. Part I: Theory and Single Error Correction. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol.39, no.1, Jan 1992, pp.8-17.
D0019	Yuan Xu, Zhao-ming Dou & Shu-pian Tang. Multipole Nonlinear Response of Metamaterials. *Proceedings of the Symposium on Photonics and Optoelectronic (SOPO)*, June 2010.	O0029	J. Petschulat, A. Chipouline, A. Tunnermann & T. Pertsch. Multipole Nonlinearity of Metamaterials. *Physical Review*, 80, 2009, American Physical Society.
D0020	Qiang Li & Yongshi Zhang. An Efficient Mining Algorithm for Top K Strongly Correlated Item Pairs. *Proceedings of the th International Conference on Internet Computing for Science and Engineering*, December 2009 , pp. 152-155.	O0030	S Roy & D K Bhattacharyya. Efficient Mining of Top-K Strongly Correlated Item Pairs using One Pass Technique. *Proceeding of the 16th International Conference on Advanced Computing and Communications*, December 2008, pp. 416-412.
D0021	Hong-Xing Zheng, Wei He and Chong Peng. D Spherical Perfectly Matched Layer for Finite-volume Domain Truncation. *Proceedings of the International Conference on Microwave and Millimeter Wave Technology* (ICMMT), 2010, pp. 113-116.	O0031	Christophe Fumeaux, Dirk Baumann & Rüdiger Vahldieck. Finite-Volume Time-Domain Analysis of a Cavity-Backed Archimedean Spiral Antenna. *IEEE Transactions on Antennas and Propagation*. Vol 54, N° 3, March 2006.
		O0274	C. Fumeaux, K. Sankaran, R. Vahldieck Spherical Perfectly Matched Absorber for Finite-volume 3-D Domain Truncation. IEEE *Transactions on Microwave Theory and Techniques*. Vol. 55, No. 12, December 2007, pp. 2773-2781.
D0022	Yucheng Zhou, Jiahe Xu, Yuanwei Jing & Georgi M. Dimirovski. Cholesky-Based Reduced-Rank Squar-Root Ensemble Kalman Filtering. *Proceedings of the American Control Conference*, July 2010 , pp 6870-6875.	O0032	J. Chandrasekar, I. S. Kim, D. S. Bernstein & A. J. Ridley Cholesky-Based Reduced-Rank Square-Root Kalman Filtering. *2008 American Control Conference*, June 11-13 2008.

D0023	Yucheng Zhou, Jiahe Xu, Yuanwei Jing, Georgi M. Dimirovski. Extended Target Tracking Using an IMM Based Nonlinear Kalman Filters. *Proceedings of the American Control Conference*, July 2010, pp. 6870-6875.	O0033	Zhiwen Zhong, Huadong Meng & Xiqin Wang Extended Target Tracking using an IMM Based Rao-Blackwellised Unscented Kalman Filter Proceedings of the 9th International Conference on Signal Processing, October 2008, pp. 2409-2412.
D0024	R. Shojaee, A. Akbari, M.A. Fard & M. Allahbakhshi. Suggesting an Agent Based Design Methodology for Condition Monitoring of High Voltage Apparatus. *Proceedings of the 10 th IEEE International Conference on Solid Dielectrics (ICSD)*, July 2010.	O0034 O0035	S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part I: Concepts, Approaches, and Technical Challenges. *IEEE Transactions on Power Systems*, Vol. 22 , No4, November 2007 , pp. 1743-1752 S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part II: Technologies, Standards, and Tools for Building Multi-agent Systems. *IEEE Transactions on Power Systems*, Vol. 22 , No.4 , November 2007 , pp. 1753-1759.
D0025	Wenjun Wang, Xiuming Li & Xiong Zeng The Application of Critical Chain Project Management Based on Max-plus Linear System. *Proceedings of the 2 nd Conference on Environmental Science and Information Application Technology (ESIAT)*, 2010, pp. 640-643.	O0036	Hirotaka Takahashi, Hiroyuki Goto & Munenori Kasahara. Application of a Critical Chain Project Management based Framework on Max-Plus Linear Systems. *International Journal of Computational Science*, vol. 3, no. 2, April 2009, pp. 117-132.
D0026	Qingshan Zhao and Guoyan Meng Design of Digital FIR Filters Using Differential Evolution Algorithm Based on Reserved Gene Proceedings of the International Conference of Information Science and Management Engineering, August 2010 , pp. 177-180.	O0037	Gang Liu, YuanXiang Li & GuoLiang He. Design of Digital FIR Filters Using Differential Evolution Algorithm Based on Reserved Genes. *Proceedings of the IEEE Congress on Evolutionary Computation*, July 2010, pp. 1-7.
D0027	ZHAO Yueling, JIN Hui, WANG Lihong & WANG Shuang. Influence Diagram Based on Rough Set Theory. Proceedings of the th Chinese Control Conference, July 2010, Beijing, China.	O0038	Chia-Hui Huang, Han-Ying Kao & Han-Lin Li. Intelligent Decision Support Based on Influence Diagrams with Rough Sets. A. An et al. (Eds.): *RSFDGrC 2007*, LNAI 4482, pp. 518–525,
D0028	Feng Dexiong and Wang Xiang. A Study on a New Pattern of Enterprise - Intelligent Enterprise. *Proceedings of the International Conference on Management and Service Science*, August 2010.	O0039	Kemal A. Delic & Umeshwar Dayal. The Rise of the Intelligent Enterprise. *ACM Ubiquity*, December 2002
D0029	Yueqian Xu. E-government and Governance In China. *Proceedings of the International Conference on Management and Service Science*, August 2010.	O0040	Ian Holliday & Ray Yep. E-Governement in China. *Public Admin. Dev*. 25, 239–249 (2005)

D0030	Hui-Xiang Xu, Su Yu, Zhi-Hong Yang, Zhi-Chao Li. A Method of Motion Capture from the Multi-view Video Sequence. *Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, Vol. 4. Aug. 2010, 492-495 .	O0041	Juergen Gall, Carsten Stoll, Edilson de Aguiar, Christian Theobalt, Bodo Rosenhahn & Hans-Peter Seidel. Motion Capture Using Joint Skeleton Tracking and Surface Estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
D0031	Majid Shakeri, and Ebrahim Hosseini. Design . GHz Wideband Bandpass Filter Using Two Section of open stub connected in shunt with microstrip lines. *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), 2010, pp. 1–5*	O0042	Pramod K. Singh, Sarbani Basu & Yeong-Her Wang. Planar Ultra-Wideband Bandpass Filter Using Edge Coupled Microstrip Lines and Stepped Impedance Open Stub. *IEEE Microwave and Wireless Components Letters*. Vol 17. N° 9 September 2007.
D0032	Du Yun-Feng and Huang Qi. Optimal Placement of Distributed Generation on a Radial Feeder. *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010, pp. 1–5	O0266	Caisheng Wang, M. Hashem Nehrir. Analytical Approaches for Optimal Placement of Distributed Generation Sources in Power Systems. *IEEE Transactions on Power Systems*, Vol. 19 No. 4, November 2004, 2068–2076.
D0033	Kai Mao. A edge to edge capacity contracting congestion pricing implementation in IP networks. *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, 2010 , pp. 1-4.	O0044 O0045	Murat Yuksel, Shivkumar Kalyanaraman & Anuj Goel. Congestion Pricing Overlaid on Edge-to-Edge Congestion Control. *IEEE International Conference on Communications 2003*, Vol. 2, 2003, 880 – 884. Murat Yuksel & Shivkumar Kalyanaraman. Distributed dynamic capacity contracting: an overlay congestion pricing framework. *Computer Communications* 26 (2003) 1484–1503
D0034	Fei Liao. Optimal Control Applied to Cholera Model. *Proceedings of the 2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010)*, Oct 2010, pp. 2662-2666	O0046	Rachael L. Miller Neilan, Elsa Schaefer, Holly Gaff, K. Renee Fister & Suzanne Lenhart. Modeling Optimal Intervention Strategies for Cholera. *Bulletin of Mathematical Biology* (2010) 72, March 2010, pp. 2004–2018.
D0035	J. Ramkumar & V.B. Kirubanand. Network Performance Management By Using Stable Algorithms. *2nd International Conference on Computer Technology and Development (ICCTD)*, 2010, pp. 692–696.	O0047	Hwisung Jung & Massoud Pedram. A Unified Framework for System-level Design: Modeling and Performance Optimization of Scalable Networking Systems. *8th International Symposium on Quality Electronic Design, 2007 (ISQED 2007)*, 2007, pp. 198–203
D0037	Lei Li & Jianhao Hu. Joint Redundant Residue Number Systems and Module Isolation for Mitigating Single Event Multiple Bit Upsets in Datapath. *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, Dec. 2010, pp. 3779-3786.	O0049	Vik Tor Goh & Mohammad Umar Siddiqi. Multiple Error Detection and Correction Based on Redundant Residue Number Systems. *IEEE Transactions on Communications*, vol.56, no.3, March 2008, pp.325-330.

		O0050	Hari Krishna, Kuo-Yu Lin & Jenn-Dong Sun. A Coding Theory Approach to Error Control in Redundant Residue Number Systems. Part I: Theory and Single Error Correction. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol.39, no.1, Jan 1992, pp.8-17.
D0038	Shenping Hu, Quangen Fang, Jinpeng Zhang, Yongtao Xi. Relative Risk Assessment Methodology in Vessel Traffic at Sea. *Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2010), pp. 1493-1498.*	O0051	Christos A. Kontovas & Harilaos N. Psaraftis. Formal Safety Assessment: A Critical Review. *Marine Technology*, Vol. 46, No. 1, January 2009, pp. 45–59
D0039	Yingchi Mao, Xiaofang Li, Yi Liang. Workload-Based Query Routing Tree Algorithm in Wireless Sensor Networks. *International Conference on Computational Intelligence and Software Engineering, (CiSE*), December 2010, pp. 1-4.	O0052	P. Andreou, A. Pamboris, D. Zeinalipour-Yazti, P.K. Chrysanthis & G. Samaras. ETC: Energy-driven Tree Construction in Wireless Sensor Networks. *Proceedings of the Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, May 2009, pp. 513-518.
D0040	Yunfeng Du and Qi Huang. Reactive Power Control Planning to Increase Voltage Stability Margin. *Proceedings of the International Conference on Computational Intelligence and Software Engineering*, December 2010.	O0053	Haifeng Liu, Licheng Jin, James D. McCalley, Ratnesh Kumar, Venkataramana Ajjarapu & Nicola Elia. Planning Reconfigurable Reactive Control for Voltage Stability Limited Power Systems. *IEEE Transactions on Power Systems*, Vol. 24, No. 2, May 2009, pp. 1029-1038.
D0041	M.L. George, L. Ngalamou. A Better Approach to Teaching Digital Logic Design to Engineering Undergraduates than Using Only One of MSISSI, Schematic Entry or HDLs. *IEEE Transactions on Education*, Early Access	O0054	Thomas Weng, Yi Zhu & Chung-Kuan Cheng. Digital Design and Programmable Logic Boards: Do Students Actually Learn More ? *38th ASEE/IEEE Frontiers in Education Conference*, October 22–25, 2008, Saratoga Springs, NY.
		O0055	Yi Zhu, Thomas Weng & Chung-Kuan Cheng Enhancing Learning Effectiveness in Digital Design Courses Through the Use of Programmable Logic Boards. IEEE Transactions on Education. Vol 52, N° 1, February 2009.
D0042	Yingchi Mao, Xiaofang Li, Yi Liang. Latency Optimized Workload-based Query Routing Tree Algorithm in Wireless Sensor Networks. *Proceedings of the nd International Conference on Information Science and Engineering (ICISE)*, December 2010 , pp. 2159-2162.	O0052	P. Andreou, A. Pamboris, D. Zeinalipour-Yazti, P.K. Chrysanthis & G. Samaras. ETC: Energy-driven Tree Construction in Wireless Sensor Networks. *Proceedings of the Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, May 2009, pp. 513-518.
D0043	Xu Yuhui & Gang Zhong. A Novel Multi Agent System Dynamic Cooperation & Automated Negotiation Model. *Proceedings of the International Conference on Multimedia Communications*, 2010, pp. 263-266.	O0056	Weijin Jiang & Xuancheng Zhou. Research on a Novel Multi-Agent System Negotiation Strategy and Model. *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing. WiCOM '08*, October 2008.

D0044	Sivaji Kaki and Tapas Chakravarty. Compact Printed Yagi Antenna for Handheld UHF RFID Reader. *Proceedings of the International Conference on Devices and Communications (ICDeCom)*	O0057	Pavel V. Nikitin & K. V. S. Rao. Compact Yagi Antenna for Handheld UHF RFID Reader. *IEEE Transactions on Microwave Theory and Techniques*, Volume 57 , Issue 5 , Part 2, 2009.
		O0268	Zhi Ning Chen, Xianming Qing and Hang Leong Chung. A Universal UHF RFID Reader Antenna. *IEEE Transactions on Microwave Theory and Techniques* Volume 57 , Issue 5 , Part 2, 2009
D0045	Li Wang, Zhiwei Ni, Yiwen Zhang, Zhang Jun Wu, Liyang Tang Pipelined-MapReduced: An Improved MapReduce Parallel Programming Model. *Proceedings of the 2011 4th International Conference on Intelligent Computation Technology and Automation*, March 2011, pp. 871-874.	O0058	Tyson Condie, Neil Conway, Peter Alvaro, Joseph M. Hellerstein, Khaled Elmeleegy & Russell Sears. MapReduce Online. *Proceedings of the th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, April 2010.
D0046	Bo Gao, Ziming Kou, Zemin Jing. Video Stabilization By Sparse and Low-rank Matrix Decomposition. *Proceedings of the International Conference on Computer and Managemen*t, May 2011, pp. 1-5.	O0059	Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu & Yi Ma. RASL: Robust Alignment by Sparse and Low-rank Decomposition for Linearly Correlated Images. *Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 763–770.
		O0060	Guofeng Zhang, Wei Hua, Xueying Qin, Yuanlong Shao & Hujun Bao. Video stabilization based on a 3D perspective camera model. *The Visual Computer*, Volume 25, Number 11, November 2009, pp. 997-1008.
D0047	Wang Yigong, Zhang Hongqi, Dai Xiangdong & Liu jiang. A Formal Framework for Security Policy Specification and Implementation. Proceedings of the International Conference on Network Computing and Information Security (NCIS'), May 2011, pp. 253-256.	O0061	Mathieu Jaume & Charles Morisset. Towards a formal specification of access control. *Proceedings of the Workshop on Foundations of Computer Security and Automated Reasoning for Security Protocol Analysis* (FCS-ARSPA'2006)
		O0062	Mathieu Jaume & Charles Morisset. A formal approach to implement access control models. *Journal of Information Assurance and Security*, 2 (2006), pp. 59-70.
D0048	Bogdan Dmytryshyn. Microfluidic Cell Culture Systems And Cellular Analysis. *Proceedings of VIIth International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, 2011, pp. 193 – 196.	O0063	Ju Hun Yeon & Je-Kyun Park. Microfluidic Cell Culture Systems for Cellular Analysis. *BIOCHIP JOURNAL*, Vol. 1, No. 1, 17-27, March 2007

D0049	Hetal Pathak. UWB Reconfigurable Microstrip Antenna for Wireless Sensor Networks Applications. *International Conference on Communications Systems and Network Technologies*, 2011, pp. 246 – 249.	O0064	Hala Elsadek , Ahmed Khidre, Adel Abdel Rahman, Abbas Omar & Hani Ragai. UWB Reconfigurable Microstrip Antenna with 3 Bit Band Rejection for Wireless Sensor Networks Applications. *14th WSEAS International Conference on Communications*, 2009, pp. 98 – 102
D0050	YuHan Gao, LimTao Liu, RuZhang Li. Model based comprehensive functional verification of Rf SoC. *Second International Conference on Digital Manufacturing and Automation* (ICDMA), October 20, 2011, pp. 728 – 731.	O0065	Yifan Wang, Stefan Joeres, Ralf Wunderlich & Stefan Heinen. Modeling Approaches for Functional Verification of RF-SoCs: Limits and Future Requirements. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. Vol 28, N° 5, May 2009.
		O0066	Yifan Wang, Christoph, Van-Meersbergen, Hans-Werner Groh & Stefan Heinen. Event Driven Analog Modeling for the Verification of PLL Frequency Synthesizers. *2009 Behavioral Modeling and Simulation Workshop (BMAS 2009)*, 2009, pp. 25 – 30.
D0051	V. Stornelli, R. Minutolo, G. Leuzzi, F. Barcio, M. Montanari A low complexity tuneable pulse generator architecture for sub-GHz UWB applications. *IEEE International Conference on Ultra-Wideband (ICUWB)*, 2011, pp. 58 – 62.	O0067	Fabio Zito, Domenico Pepe, & Domenico Zito. UWB CMOS Monocycle Pulse Generator. *IEEE Transactions on Circuits and Systems. I Regular Papers*. Vol 57, N° 10, October 2010.
D0052	Shan Wang, Xiaoxi Zhang, Huicheng Feng. Map-Based Location with Base Stations. *IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC*), September 2011.	O0068	Marco Anisetti, Claudio A. Ardagna, Valerio Bellandi, Ernesto Damiani & Salvatore Reale Map-Based Location and Tracking in Multipath Outdoor Mobile Networks. IEEE Transactions on Wireless Communications. Vol 10, N° 3, March 2011.
D0053	Vimal Kant Tyagi, Sunil Kumar Chowdhary, Naveen Garg. Authentication Using Graphical Password To Upgrade Security & Memorability. *Proceedings of IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, October 2011.	O0069	Haichang Gao, Xuewu Guo, Xiaoping Chen, Liming Wang, and Xiyang Liu. YAGP: Yet Another Graphical Password. *Proceedings of the 2008 13th Asia-Pacific Computer Systems Architecture Conference*, (ACSAC), August 2008, pp. 121—129.
D0054	Mei Chen, Liang Guo. A Correction of LSV Herding Measure. *Proceedings of the Fourth International Conference on Business Intelligence and Financial Engineering*, 2011, pp. 362-366.	O0070	Raphaëlle Bellando. Measuring herding intensity: a hard task. *Social Science Research Network*-June 2010.
D0055	Mahdi Jalili-Kharaajoo and Hossein Rouhani. Robust Nonlinear Control Applied to Traction Control of Electric Vehicles. *Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems*,. ICECS 2003, December 2003.	O0071	Remus Pusca, Youcef Ait-Amirat, Alain Berthon & Jean Marie Kauffmann Modeling and Simulation of a Traction Control Algorithm for an Electric Vehicle with Four Separate Wheel Drives Proceedings of the 2002 IEEE 56th Vehicular Technology Conference, VTC 2002-Fall, pp. 1671-1675

		O0272	Feijun Song, Samuel M. Smith. A Comparison of Sliding Mode Fuzzy Controller and Fuzzy Sliding Mode Controller. *Proceedings of the 19th International Conference of the North American Fuzzy Information Processing Society*, 2000. NAFIPS, July 2000, pp 480-484.
D0056	Jiayue Tang, Jie Wang. Adaptive Modulation for Fading Channels. *Proceedings of the 2000 IEEE Asia-Pacific Conference on Circuits and Systems*, 2000, (APCCAS 2000), pp.58-61, December 2000.	O0072	Nilo Casimiro Ericsson. Adaptive Modulation and Scheduling of IP traffic over Fading Channels. *Proceedings of the 1999 Global Telecommunications Conference*, 1999. GLOBECOM '9, vol.5, pp. 2668-2672
D0057	Jiang Guo. Software Components Adaptive Integration. *Proceedings of Eighth Annual IEEE International Conference and Workshop on the Engineering of Computer Based Systems* (ECBS), 2001	O0073	John Penix & Perry Alexander. Toward Automated Component Adaptation. *Proceedings of the Ninth International Conference on Software Engineering and Knowledge Engineering* (SEKE-97), pp 535.542.
		O0074	John Penix. REBOUND: A Framework for Automated Component Adaptation. *Proceedings of the 9th Annual Workshop on Software Reuse (WISR-99)*, 1999.
D0058	Tripathy A. K. & Singh A. K. An Efficient Method of Eliminating Noisy Information in Web Pages for Data Mining *Fourth International Conference on Computer and Information Technology (CIT'04)* September 14-16, 2004, Wuhan, China, pp. 978-985.	O0075	Lan Yi, Bing Liu & Xiaoli Li . Eliminating Noisy Information in Web Pages. *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2003)*, Washington, DC, USA, August 24-27, 2003.
D0059	H.Z Tameem & B.V Mehta. Solid Modeling of Human Vocal Tract Using Magnetic Resonance Imaging and Acoustic Pharyngometer. *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004. IEMBS '04. Volume 2, September 2004	O0076	W. Tecumseh Fitch & Jay Giedd. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Acoustical Society of America*, September 1999, pp. 1511-1522
D0060	Jiang Guo, Yuehong Liao, Jeff Gray & Barrett Bryant. Using Connectors to Integrate Software Components. *Proceedings of 12th IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS'05)*	O0073	John Penix & Perry Alexander. Toward Automated Component Adaptation. *Proceedings of the Ninth International Conference on Software Engineering and Knowledge Engineering* (SEKE-97), pp 535.542.
		O0074	John Penix. REBOUND: A Framework for Automated Component Adaptation. *Proceedings of the 9th Annual Workshop on Software Reuse (WISR-99)*, 1999.

D0061	R. Bhattacharya, P. Venkateswaran, S.K. Sanyal, and R. Nandi. Intelligent Queue Management Protocol for TCP Traffic over 3G Links. *Proceedings of the International Conference on Personal Wireless Communications 2005*, pp. 496-499.	O0077	Mats Sågfors, Reiner Ludwig, Michael Meyer & Janne Peisa. Queue Management for TCP Traffic over 3G Links. *Proceedings of the Wireless Communications and Networking Conference* 2003, vol. 3, pp 1663-1668
D0062	L. Gatani, G. Lo Re, S. Gaglio An Adaptive Routing Protocol for Ad Hoc Peer-to-peer Networks Proceedings of the Sixth International Symposium on World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. pp. 44-50, 13-16 June 2005	O0078	Filippo Menczer, Ruj Akavipat & Le-Shin Wu.. 6S: Distributing crawling and searching across Web peers. *CIKM 2004* Washington, D.C. USA
D0063	Alimujiang Yiming and Toshio Eisaka. An Ethernet Protocol for Real-time Communications. *Proceedings of the SICE Annual Conference in Sapporo*, August 2004	O0079	Hoai Hoang, Magnus Jonsson, Ulrik Hagstrom & Anders Kallerdahl. Witched Real-Time Ethernet With Earliest Deadline First Sscheduling-Protocols, Traffic Handling and Simulation Analysis. *Parallel and Distributed Computing Practices*, 2002, vol 5, 1, pp 105-115.
D0064	Abed Elhamid Lawabni & Ahmed H. Tewfik. Resource Management and Quality Adaptation in Distributed Multimedia Networks. *Proceedings of the 10th IEEE Symposium on Computers and Communications (ISCC 2005)*, pp 604-610.	O0080	Md Mostofa Akbar, Eric G. Manning, Gholamali C. Shoja & Shahadat Khan Heuristic Solutions for the Multiple-Choice Multi-Dimension Knapsack Problem. V.N. Alexandrov et al. (Eds.): *ICCS 2001, LNCS 2074*, pp. 659−668.
		O0081	Md Mostofa Akbar, M. Sohel Rahman, M. Kaykobad, E.G. Manning & G.C. Shoja. Solving the Multidimensional Multiple-choice Knapsack Problem by constructing convex hulls. *Computers & Operations Research* 33 (2006) 1259–1273
		O0082	Lei Chen, Shahadat Khan, Kin F. Li, and Eric G. Manning. Building an Adaptive Multimedia System using the Utility Model Parallel and Distributed Processing, *Lecture Notes in Computer Science*, 1586, Springer Verlag, pp 289-298.
		O0028	S. Khan, K. F. Li, E. G. Manning, M. M. Akbar. Solving the Knapsack Problem for Adaptive Multimedia Systems. *Studia Informatica*, 2002
D0065	Kaewmannee, P.; Ormondroyd, R.F.; Walters, C.R. Resilient adaptive wavelet packet modulation scheme for use in time and frequency selective channels using a best tree search algorithm. *Proceedings of the IEEE Military Communications Conference 2004 (MILCOM 2004*), vol. 3, 31 Oct.-3 Nov. 2004, pp. 1566 - 1571.	O0083	D. Daly , C. Heneghan, A. Fagan & Martin Vetterli. Optimal Wavelet Packet Modulation under Finite Complexity Constraint. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2002*, (ICASSP '02), vol. 3, 13-17 May 2002, pp. :III-2789

D0066	Haipeng Wang, Peng Tian, and Ping Jin. Electricity Consumption and Economic Growth in China. *Proceedings of the 2005 International Conference on Services Systems and Services Management*, vol. 2, pp 1331-1334.	O0084	Alice Shiu & Pun-Lee Lam. Electricity consumption and economic growth in China. *Energy Policy*. Vol. 32, no. 1, Jan 2004, Elsevier, pp 47-54.
D0067	L. Gatani, G. Lo Re, L. Noto. Efficient Query Routing in Peer-to-peer Networks. *Proceedings of the Third International Conference on Information Technology: Research and Education*, 2005. ITRE 2005. pp. 393-397, 27-30..	O0078	Filippo Menczer, Ruj Akavipat & Le-Shin Wu.. 6S: Distributing crawling and searching across Web peers. *CIKM 2004* Washington, D.C. USA
D0068	Konstantinos Fotiades and John Seimenis. Fast Lane Changing Algorithm for Intelligent Vehicle Highway Systems Using Clothoidal Theory and Bezier Points. *Proceedings of the Intelligent Vehicles Symposium*. 2005, pp. 73-77	O0085	N.Montés & J.Tornero. Lane changing using s-series clothoidal approximation and dual-rate based on Bezier points to controlling vehicle. *WSEAS Transactions on Circuits and Systems*, vol 3, pp. 2285-2290, December 2004.
D0069	L. Gatani, G. Lo Re, A. Urso, S. Gaglio. Reinforcement Learning for P2P Searching. *Proceedings of the Seventh International Workshop on Computer Architecture for Machine Perception*, 2005. CAMP 2005. pp. 303-308, 4-6 July 2005.	O0078	Filippo Menczer, Ruj Akavipat & Le-Shin Wu.. 6S: Distributing crawling and searching across Web peers. *CIKM 2004* Washington, D.C. USA
D0070	E. I. Kalantzis, P.I. Dallas, and B. S. Sharif. A Novel Stopping Criterion for Turbo Decoding. *Proceedings of the 2005 IEEE Vehicular Technology Conference* (VTC), pp. 24-37.	O0086 O0087	Yufei Wu, Brian D. Woerner & William J. Ebel. A Simple Stopping Criterion for Turbo Decoding. IEEE Communications Letters, Vol 4, N° 8, August 2000. Rose Y. Shao, Shu Lin & Marc P. C. Fossorier. Two Simple Stopping Criteria for Turbo Decoding. IEEE Transactions on Communications, Vol 47, N° 8, August 1999, pp. 117-1120.
D0071	L. Gatani, G. Lo Re, S. Gaglio. An Adaptive Routing Mechanism for P2P Resource Discovery. *Proceedings of the IEEE International Symposium on Cluster Computing and the Grid*, 2005. CCGrid 2005. pp. 205-212 Vol. 1, 9-12 May 2005.	O0078	Filippo Menczer, Ruj Akavipat & Le-Shin Wu.. 6S: Distributing crawling and searching across Web peers. *CIKM 2004* Washington, D.C. USA.
D0072	M.B.K. Ghembaza, Y. Amirat, K. Djouani, B. Daachi. Deformation Model of Soft Tissue for Abdominal Aorta and Aneurysm. *2005 Proceedings of the International Conference on Systems, Man and Cybernetics*, vol. 4, pp 3574-3579.	O0088	Jean-Marc Schwartz, Marc Denninger, Denis Rancourt, Christian Moisan & Denis Laurendeau. Modelling liver tissue properties using a non-linear visco-elastic model for surgery simulation. *Medical Image Analysis* 9 (2005) 103–112
D0073	Abed Elhamid Lawabni & Ahmed H. Tewfik. Resource Management and Knapsack Formulation in Distributed Multimedia Networks. *IEEE Global Telecommunications Conference* (Globecom 2005), pp. 413-418.	O0080	Md Mostofa Akbar, Eric G. Manning, Gholamali C. Shoja & Shahadat Khan Heuristic Solutions for the Multiple-Choice Multi-Dimension Knapsack Problem. V.N. Alexandrov et al. (Eds.): *ICCS 2001, LNCS 2074*, pp. 659−668.

		O0081	Md Mostofa Akbar, M. Sohel Rahman, M. Kaykobad, E.G. Manning & G.C. Shoja. Solving the Multidimensional Multiple-choice Knapsack Problem by constructing convex hulls. *Computers & Operations Research* 33 (2006) 1259–1273
		O0082	Lei Chen, Shahadat Khan, Kin F. Li, and Eric G. Manning. Building an Adaptive Multimedia System using the Utility Model Parallel and Distributed Processing, *Lecture Notes in Computer Science*, 1586, Springer Verlag, pp 289-298.
		O0028	S. Khan, K. F. Li, E. G. Manning, M. M. Akbar. Solving the Knapsack Problem for Adaptive Multimedia Systems. *Studia Informatica*, 2002
D0074	Thida Win, Hninn Mar Aung, Ni Lar Thein. An MDA Based Approach for Facilitating Representation of Semantic Web Service Technology. *Proceedings of the 6th Asia-Pacific Symposium on Information and Telecommunication Technologies*, 2005 (APSITT 2005), Nov. 2005, pp. 260-265.	O0089	John T. E. Timm & Gerald C. Gannod. A Model-Driven Approach for Specifying Semantic Web Services. *Proceedings of the 2005 IEEE International Conference on Web Services* (ICWS 2005), vol.1, July 2005, pp. 313- 320.
		O0273	Gerald C. Gannod & John T. E. Timm. An MDA-based Approach for Facilitating Adoption of Semantic Web Service Technology. *Proceedings of the IEEE EDOC Workshop on Model-Driven Semantic Web (MDSW04*), Sept. 2004, pp. 654-675.
D0075	V. Zamani Farahani, Sh. Jadid, H. Mohammadi Bidhendi. Forward Energy Contracts for Electricity Markets under Market. Power and Network Uncertainty. *Proceedings of the 2006 IEEE Power India Conference*.	O0090	Rajnish Kamat & Shmuel S. Oren. Two-settlement Systems for Electricity Markets under Network Uncertainty and Market Power. *Journal of Regulatory Economic*s, 25-1, 2004, 5-37
D0076	S. Muhammad Siddique & Muhammad Amir. GSM Security Issues and Challenges. *Proceedings of the Seventh IEEE International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* (SNPD'06), pp.413-418.	O0091	Jeremy Quirke. Security in the GSM system. *Ausmobile*, May 2004
D0077	A. Falahati, M. Mashreghi, M. Tafaroji. Security Enhancement in CDMA with a Hidden Direct Sequence Spread Spectrum System. *2nd Information and Communication Technologies (ICTTA 2006)*, 2006, pp. 2524 – 2529.	O0092	Jian Ren & Tongtong Li. CDMA Physical Layer Built-in Security Enhancement. *2003 IEEE 58th Vehicular Technology Conference (VTC 2003)*, Vol. 3, 2003, pp. 2157-2161.

D0078	K.W. Yeom, J.H. Lee & J.H. Park. Biologically Inspired Evolutionary Agent Systems in Dynamic Environments. *Proceedings of the International Conference on Computational Intelligence and Security*, 2006.	O0105	Josh C. Bongard & Rolf Pfeifer. Evolving Complete Agents using Artificial Ontogeny. *Morpho-functional Machines: The New Species (Designing Embodied Intelligence)*, Springer-Verlag, pp. 237-258.
D0079	Yu Renjun & Tan Eng Chong. Spectrogram Analysis of Genome Small Patterns Using Pseudo Smoothed Wigner-Ville Distribution. *Proceedings of the 5th International Conference on Information, Communications & Signal Processing*, December 2005.	O0094	David Sussillo, Anshul Kundaje & Dimitris Anastassiou. Spectrogram Analysis of Genomes. *EURASIP Journal on Applied Signal Processing* 2004:1, 29–42
D0080	Changfeng Zhou, Yan Liu, Yuejin Tan & Liangcai Liao. Dynamic Vehicle Routing and Scheduling with Variable Travel Times in Intelligent Transportation System. *Proceedings of the 6th World Congress on Intelligent Control and Automation*, June 2006, pp 8707-8711.	O0095	Eiichi Taniguchi & Hiroshi Shimamoto. Intelligent transportation system based dynamic vehicle routing and scheduling with variable travel times. *Transportation Research*, Part C 12 (2004), 235–250
D0081	Xin Jin, Rongfang Bie & X.Z. Gao. An Artificial Immune Recognition System-based Approach to Software Engineering Management: with Software Metrics Selection. *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications*, 2006, pp. 523-528.	O0096	Jason Brownlee. *Artificial Immune Recognition System (AIRS). A Review and Analysis*. Swinburne University of Technology, January 2005.
D0082	Jue Wang, Changjun Hu, Jianxin Lai, Yudi Zhao & Suqin Zhang. Multi-paradigm and Multi-grain Parallel Execution Model Based on SMP-Cluster. *Proceedings of the IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing*, 2006 (JVA '06) pp.266-272.	O0097	Gabriele Jost, Haoqiang Jin, Dieter an Mey & Ferhat F. Hata. Comparing the OpenMP, MPI, and Hybrid Programming Paradigms on an SMP Cluster. *Fifth European Workshop on OpenMP* (EWOMP03) in Aachen, Germany, September 2003.
D0083	Xiao-Li Dong, Cheng-Kui Gu, Zheng-Ou Wang. A Local Segmented Dynamic Time Warping Distance Measure Algorithm for Time Series Data Mining. *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*, 2006.	O0098	Eamonn J. Keogh and Michael J. Pazzani. Scaling up Dynamic Time Warping to Massive Datasets. *Third European Conference on Principles of Data Mining and Knowledge Discovery*, Springer-Verlag, 1999.
D0084	Hua Chen, Qin-Ming He, Jian-Fei Qian. Cognitive Overhead Reducing Based on Collaborative Filtering in Project Management. *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*, 2006.	O0099	Ken Goldberg, Theresa Roeder, Dhruv Gupta & Chris Perkins. Eigentaste: A Constant Time Collaborative Filtering Algorithm. *Information Retrieval*, Vol. 4, No. 2, 2001, pp. 133-151.
D0085	Tong-Cheng Huang and You-Dong Ding. Generic Object Recognition Via Integrating Distinct Features with SVM. *Proceedings of 2006 International Conference on Machine Learning and Cybernetics*, pp 3897-3902.	O0100	Hongying Meng, David R. Hardoon, John Shawe-Taylor & Sandor Szedmak. Generic object recognition by combining distinct features in machine learning. *Proceedings of the 17th Annual Symposium on Electronic Imaging*, January 2005.

D0086	Hongxia Xie and Zhengyun Hui. An Intrusion Detection Architecture for Ad hoc Network Based on Artificial Immune System. *Proceedings of the 7th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'06)*, Taipei, Taiwan, Dec. 2006.	O0101	John Hall. Comparing an Immunological and a Rule-Based Intrusion Detection Method. *CSDS Technical Report*, University of Idaho 2002.
D0087	Jian-Hua Song, Fan Hong, and Yu Zhang. Effective Filtering Scheme against RREQ Flooding Attack in Mobile Ad Hoc Networks. *Proceedings of the Seventh IEEE International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)* in 2006.	O0102	Sugata Sanyal, Ajith Abraham, Dhaval Gada, Rajat Gogri, Punit Rathod, Zalak Dedhia & Nirali Mody. Security Scheme for Distributed DoS in Mobile Ad Hoc Networks. *Sixth International Workshop on Distributed Computing (IWDC)*, vol. 3326, LNCS. Springer, 2004.
D0088	M. Hatti, M. Tioursi, and W. Nouibat. A Q-Newton Method Neural Network Model for PEM Fuel Cells. *Proceedings of the 4th IEEE International Conference on Industrial Informatics* (INDIN), August 2006.	O0103	S. Jemeï, D. Hissel, M.C. Péra & J.M. Kauffmann. On-board fuel cell power supply modeling on the basis of neural network methodology. *Journal of Power Sources* 124 (2003), 479–486.
D0089	E. Ardizzone, L. Gatani, M. La Cascia, G. Lo Re, M. Ortolani. Enhanced P2P Services Providing Multimedia Content. *Proceedings of the Eighth IEEE International Symposium on Multimedia*, 2006. ISM 2006. pp.637-646, Dec. 2006.	O0076	W. Tecumseh Fitch & Jay Giedd. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Acoustical Society of America*, September 1999, pp. 1511-1522
D0090	Dong Li and Wenqiang Yu. A New Predictive Mechanism Based on Artificial Neural Network. *Proceedings of the International Conference on Computational Intelligence and Security*, Nov. 2006 pp. 333-338.	O0104	Alaknantha Eswaradass, Xian-He Sun & Ming Wu. A Neural Network Based Predictive Mechanism for Available Bandwidth. *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium*, April 2005.
D0091	K.W. Yeom and J.H. Park. Biologically Inspired Evolutionary Agent Systems in Dynamic Environments. *Proceedings of the IEEE Congress on Evolutionary Computation*, 2006. CEC 2006, pp.386-390.	O0093	Iacopo Carreras, Francesco De Pellegrini, Daniele Miorandi & Hagen Woesner. Service Evolution in a Nomadic Wireless Environment. *Lecture Notes in Computer Science*, 2006, Springer, pp. 92-106.
D0092	Irahis Rodriguez and Roberto Alves. Bearing Damage Detection of the Induction Motors using Current Analysis. *Proceedings of the IEEE/PES Transmission and Distribution Conference and Exposition*, 2006	O0106	Martin Blödt, Pierre Granjon, Bertrand Raison & Gilles Rostaing. Models for Bearing Damage Detection in Induction Motors Using Stator Current Monitoring. *International Symposium on Industrial Electronics*, 2004 (ISIE'04) Ajaccio, France
D0093	Liu Wenke, Jin Liang, Cheng Baowei, Hu Junjie. Joining Power Control and Smart Antenna Using Kalman Algorithm for CDMA Communication Systems. *Proceedings of the 2006 International Conference on Signal Processing*, November 2006	O0107	Mohammed Elmusrati & Heikki Koivo. Joining Power Control and Smart Antenna Using Kalman Algorithm for CDMA Communication Systems. *Proceedings of the 2002 General Assembly of the International Union of Radio Science*, August 2002.

D0094	I.M. Culbert and W. Rhodes. Using Current Signature Analysis Technology to Reliably Detect Cage Winding Defects in Squirrel-Cage Induction Motors. *IEEE Transactions on Industry Applications*, Vol. 43, No. 2, March/April 2007	O0108	Mark Fenger, Ian Culbert, Greg Stone & William T. Thomson. Development of a Tool to Detect Faults in Induction Motors via Current Signature Analysis. *Electric Motor Predictive Maintenance Conference*, August 2002
D0095	N.B. Muthuselvan, P. Somasundaram, and Subhransu Sekhar Dash. Security Enhancement of Optimal Power Flow using Genetic Algorithm. *Proceedings of the International Conference on Power Electronics, Drives and Energy Systems*, December 2006	O0109	D. Devaraj & B. Yegnanarayana. Genetic-algorithm-based optimal power flow for security enhancement. *IEEE Proceedings-Generation, Transmission and Distribution*, Vol 152, No 6, November 2005, p p. 899-905
D0096	Lingyun Jiang, Guangzeng Feng. A MAC Aware Cross-Layer Routing Approach for Wireless Mesh Network. *International Conference on Wireless Communications, Networking and Mobile Computing* (WICCOM), 2006	O0110	Golnaz Karbaschi & Anne Fladenmuller. A Link-Quality and Congestion-aware Cross layer Metric for Multi-Hop Wireless Routing. *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference (MASS)*, Nov. 2005
D0097	M. Hatti, M. Tioursi, and W. Nouibat. A Q-Newton Method Neural Network Model for PEM Fuel Cells. *Proceedings of the First International Symposium on Environment Identities and Mediterranean Area*, 2006. ISEIMA '06, July 2006	O0103	S. Jemeï, D. Hissel, M.C. Péra & J.M. Kauffmann. On-board fuel cell power supply modeling on the basis of neural network methodology. *Journal of Power Sources* 124 (2003), 479–486.
D0098	D. Kouvatsos, and S.A. Assi. On the Analysis of Queues with Long Range Dependent Traffic: An Extended Maximum Entropy Approach. *Proceedings of the 3rd Euro NGI Conference on Next Generation Internet Networks*, May 2007, pp. 226-233	O0111	Karmeshu & Shachi Sharma. Long Tail Behavior of Queue Lengths in Broadband Networks: Tsallis Entropy Framework. July 2005 : http://arxiv.org/abs/1012.2464.
D0099	Mohd Fadlee A. Rasid, Raja Syamsul Azmir Raja Abdullah, M. Hossein Fotouhi Ghazvini and Maryam Vahabi. Energy Optimization with Multi-level Clustering Algorithm for Wireless Sensor Networks. *Proceedings of the International Conference on Wireless and Optical Communications Networks*, 2007. WOCN '07, 2-4 July 2007	O0112	Frank Comeau, Shyamala C. Sivakumar, William Robertson & William J. Phillips. Energy Conserving Architectures and Algorithms for Wireless Sensor Networks. *Proceedings of the 39th Hawaii International Conference on System Sciences* - 2006
D0100	Ahmed El Oualkadi, Luc Vandendrope, Denis Flandre. System-Level Analysis of O-QPSK Transceiver for 2.4-GHZ Band IEEE 802.15.4 Zigbee Standard. *Proceedings of the 14th International Conference on Mixed Design of Integrated Circuits and Systems*, 2007. MIXDES '07, June 2007	O0113	Nam-Jin Oh & Sang-Gug Lee. Building a 2.4-GHZ Radio Transceiver Using IEEE 802.15.4. *IEEE Circuits & Devices Magazine*. November-December 2005, pp. 43-51.
D0101	Min Gan, Guoying Yue, Hongfa Wang. Network Protocols Analyzing by an Environmental Bisimulation Method. *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, July 2007, pp 363-367.	O0114	Y.H. LU, Y.G. Gut, X.R. Ched & Y. Fui. Analyzing Security Protocols by a Bisimulation Method Based on Environmental Knowledge. *Proceeding of 2005 International Conference on Communications, Circuits and Systems*, May 2005, pp. 79-83

D0102	A. Govardhan, S. Charles, S.F. Sultana. Interference Detection in Optical CDMA Networks. *Proceedings of the International Conference on Advanced Computing and Communications, 2006*. (ADCOM 2006), pp.393-396, Dec. 2006	O0115	Purushotham Kamath, Joseph D. Touch & Joseph A. Bannister. Algorithms for Interference Sensing in Optical CDMA Networks. *Proceedings of the IEEE International Conference on Communications*, 2004, vol.3, pp. 1720-1724, June 2004
D0103	Abu Baker, S Soumik Ghosh, Ashok Kumar, and Magdy Bayoumi. LDPC Decoder: A Cognitive Radio Perspective for Next Generation (XG) Communication. *IEEE Circuits and Systems Magazine*, Third Quarter 2007, Vol. 7, Issue 3, pp. 24-37.	O0116 O0117	Chia-han Lee & Wayne Wolf. Energy/Power Estimation for LDPC Decoders in Software Radio Systems. *Proceedings of the IEEE Workshop on Signal Processing Systems Design and Implementation*, November 2005, pp. 48-53. Ian F. Akyildiz, Won-Yeol Lee, Mehmet C. Vuran & Shantidev Mohanty. NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer Networks Journal*, 50 (2006), pp. 2127–2159.
D0104	Changjun Hu, Jilin Zhang, Jue Wang, Jianjiang Li, and Liang Ding. A New Parallel Gauss-Seidel Method by Iteration Space Alternate Tiling. *Proceedings of the 16th International Conference on Parallel Architecture and Compilation Techniques*, 2007 (PACT 2007) pp.410-	O0118	Rohallah Tavakoli & Parviz Davami. A new parallel Gauss–Seidel method based on alternating group explicit method and domain decomposition method. *Applied Mathematics and Computation*, 188 (2007), pp. 713–719.
D0105	Weiwei Li, Huixian Huang, Dongbo Zhang, Hongzhong Tang and Chenhao Wang. A Color Image Segmentation Method Based on Automatic Seeded Region Growing. *Proceedings of the 2007 IEEE International Conference on Automation and Logistics*, pp. 1925-1927	O0119	Frank Y. Shih & Shouxian Cheng. Automatic seeded region growing for color image segmentation. *Image and Vision Computing*, 23, (2005), pp. 877–886.
D0106	Shi Biao, Tang Zhen-jun, Li Xian-hua, Jiang Shi-jin, and Yu Xin-hua. A Simple Subspace-Based Blind Channel Estimation Technique for OFDM Systems. *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing*, 2007. WiCom 2007, pp.236-241, Sept. 2007	O0120	Feifei Gao & A. Nallanathan. Blind Channel Estimation for MIMO OFDM Systems via Nonredundant Linear Precoding. IEEE Transactions on Signal Processing. Vol 55, N° 2, February 2007.
D0107	Zhang Qijun, Xu Qinghua, Zhu Wei. New EVM Calculation Method for Broadband Modulated Signals and Simulation. *Proceedings of the Eighth International Conference on Electronic Measurement and Instruments*, July 2007.	O0121	Michael D. McKinley, Kate A. Remley, Maciej Myslinski, J. Stevenson Kenney, Dominique Schreurs, Bart Nauwelaers. EVM Calculation for Broadband Modulated Signals. *64th ARFTG Conf. Dig*., Orlando, FL, Dec 2004, p 45-52.
D0108	M.A. Rahman and A. El Saddik. Modified Syntactic Method to Recognize Bengali Handwritten Characters. *IEEE Transactions on Instrumentation and Measurement*, Vol 56, No. 6, December 2007, pp. 2623-2632.	O0122	Mohammad Badiul Islam, Mollah Masum Billah Azadi, Md. Abdur Rahman & M. M. A. Hashem. Bengali Handwritten Character Recognition Using Modified Syntactic Method. *Proceedings of 2nd National Conference on Computer Processing of Bangla*, 2005, pp. 264-275.

D0109	Yu Zhang. Achieving Flexible Task Delegation in Role-Based Agent Teams. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2007, pp.3801-3806, 7-10 Oct. 2007	O0123	Sen Cao. *Role-base and Agent-oriented Teamwork Modeling*. Thesis, Texas A&M University, August 2005
D0110	Christina B. Vilakazi, Tshilidzi Marwala. Agents and Multi-agents Systems and Application to Condition Monitoring. *Proceedings of the 2007 International Conference on Systems, Man and Cybernetic*s 2007, pp. 644-649.	O0124	Stephen D. J. McArthur, Scott M. Strachan & Gordon Jahn. The Design of a Multi-Agent Transformer Condition Monitoring System. *IEEE Transactions on Power Systems*, Vol. 19, N° 4, November 2004, pp. 1845-1852.
D0111	Jing Nie, Liaoyuan Zeng, Jiangchuan Wen. A Bandwidth Based Adaptive Fuzzy Logic Handoff in IEEE 802.16 and IEEE 802.11 Hybrid Netwo. *Proceedings of the International Conference on Convergence Information Technology*, 2007, pp.24-29,	O0125	Amir Majles & Babak H. Khalaj. An Adaptive Fuzzy Logic Based Handoff Algorithm for Interworking Between WLANs and Mobile Networks. *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 2002. vol.5, pp. 2446- 2451,
D0112	Min Zhang; Renguo Jiang; Haiqing Hu. Security in E-Commerce and the Economics of Immediate Satisfaction. *Proceedings of the International Conference on Computational Intelligence and Security Workshops*, 2007. CISW 2007. Dec. 2007, pp.588-591.	O0126	Alessandro Acquisti. Privacy in Electronic Commerce and the Economics of Immediate Gratification. *Proceedings of the 5th ACM Conference on Electronic Commerce, Association for Computing Machinery*, New York, 2004.
D0113	Pabitra Mohan Khilar and Sudipta Mahapatra. A Dynamic Distributed Diagnosis Algorithm for an Arbitrary Network Topology with Unreliable Nodes and Links. *Proceedings of the International Conference on Advanced Computing and Communications*, 2007. ADCOM, Dec. 2007, pp. 125-130.	O0127	Elias Procopio Duarte Jr.& Andrea Weber. A Distributed Network Connectivity Algorithm. *The Sixth International Symposium on Autonomous Decentralized Systems*, 2003. ISADS 2003, April 2003, pp. 285- 292
D0114	P. Selvi, and N.P. Gopalan. Sentence Similarity Computation Based on WordNet and Corpus Statistics. *Proceedings of the International Conference on Conference on Computational Intelligence and Multimedia Applications*, 2007, vol.1, pp.9-14.	O0128	Yuhua Li, Zuhair Bandar, David McLean & James O'Shea. A Method for Measuring Sentence Similarity and its Application to Conversational Agents. *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, pp. 820–825.
		O0129	Hoa A. Nguyen & Hisham Al-Mubaid. A Combination-based Semantic Similarity Measure Using Multiple Information Sources. *Proceedings of the 2006 IEEE International Conference on Information Reuse and Integration*, pp.617-621
		O0130	Yuhua Li, David McLean, Zuhair A. Bandar, James D. O'Shea & Keeley Crockett. Sentence Similarity Based on Semantic Nets and Corpus Statistics. *IEEE Transactions on Knowledge and Data Engineering*, 2006, vol.18, no.8, pp.1138-1150

		O0131	Aminul Islam & Diana Inkpen. Semantic Similarity of Short Texts. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, Bulgaria, September 2007
D0115	R. Shantha Selva Kumari, A. Thilagamanimala, and V. Sadasivam. ECG Signal Interferences Removal Using Wavelet Based CSTD Technique. *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications*, vol. 1, December 2007, pp. 530-534.	O0132	Elvir Causevic, Robert E. Morley, M. Victor Wickerhauser & Arnaud E. Jacquin. Fast Wavelet Estimation of Weak Biosignals. *IEEE Transactions on Biomedical Engineering*, vol.52, no.6, June 2005, pp. 1021-1032.
D0116	E. Ardizzone, L. Gatani, M. La Cascia, G. Lo Re, M. Ortolani. Distributed Multimedia Digital Libraries on Peer-to-peer Networks. *Proceedings of the 14th International Conference Image Analysis and Processing Workshops*, 2007. ICIAPW 2007. pp.77-82.	O0078	Filippo Menczer, Ruj Akavipat & Le-Shin Wu.. 6S: Distributing crawling and searching across Web peers. *CIKM 2004* Washington, D.C. USA
D0117	Abdelhafidi, Z.; Djoudi, M.; Yagoubi, M.B. Simulation Based Study of Communication Induced Checkpointing Protocols Ensuring the RDT Property. *Proceedings of the 4th International Conference on Innovations in Information Technology*, Nov. 2007, pp.655-659.	O0134	Gustavo M. D. Vieira & Luiz E. Buzato. Distributed Checkpointing: Analysis and Benchmarks. *Proceedings of the 24th Brazilian Symposium on Computer Networks*, May 2006.
D0118	M. Hossein Fotouhi Ghazvini, Maryam Vahabi, Mohd Fadlee A. Rasid and Raja Syamsul Azmir Raja Abdullah. Optimizing Energy Consumption in Hierarchical Clustering Algorithm for Wireless Sensor Networks. *Proceedings of the 2007 IEEE International Conference on Telecommunications and Malaysia International Conference on Communications*. pp. 550-555.	O0112	Frank Comeau, Shyamala C. Sivakumar, William Robertson & William J. Phillips. Energy Conserving Architectures and Algorithms for Wireless Sensor Networks. *Proceedings of the 39th Hawaii International Conference on System Sciences* - 2006
D0119	Ahmed El Oualkadi, Denis Flandre HDL System-Level Design of O-QPSK Receiver for 2.4-GHz Band IEEE 802.15.4 *Proceedings of the International Conference on Design & Technology of Integrated Systems in Nanoscale Era*, 2007.	O0113	Nam-Jin Oh & Sang-Gug Lee. Building a 2.4-GHZ Radio Transceiver Using IEEE 802.15.4. *IEEE Circuits & Devices Magazine*. November-December 2005, pp. 43-51.
D0120	R. Saeedfar and P. Rabanifar. An Extended Reliability Model of Unified Power Flow Controller. *Proceedings of the 42nd International Universities Power Engineering Conference*, 2007.	O0136	F. Aminifar, M. Fotuhi-Firuzabad & R. Billinton. Extended reliability model of a unified power flow controller. Generation, Transmission & Distribution, *IET* , vol.1, no.6, Nov. 2007, pp.896-903.
D0121	T. J. Hammons. International Practices in Distributed Generation Developments Worldwide. *Proceedings of the 42nd International Universities Power Engineering Conference*, 2007 (UPEC), pp.885-894, 4-6 Sept. 2007	O0137	Kwang Y. Lee & Se Ho Kim. Progress in Distributed Generation in Korea. *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, 24-28 June 2007

		O0138	Nikos. D. Hatziargyriou, Zoe Vrontisi & Antonis G. Tsikalakis. The effect of island interconnections on the increase of Wind Power penetration in the Greek System. *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, 24-28 June 2007
D0122	T. J. Hammons. Dispersed Generation and its Impact in Europe on Power System Structure and Secure Power System Operation. *Proceedings of the 42nd International Universities Power Engineering Conference*, 2007 (UPEC), pp. 930-937.	O0139	J. Driesen, Member, IEEE, G. Deconinck, W. D'haeseleer & R. Belmans. Active User Participation in Energy Markets Through Activation of Distributed Energy Resources. *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, 24-28 June 2007
		O0140	B. Meyer. Distributed Generation: towards an effective contribution to power system security. *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, 24-28 June 2007
		O0141	Antje G. Orths, Peter B. Eriksen & Vladislav Akhmatov. Planning under Uncertainty. Securing Reliable Electricity Supply in Liberalized Energy Markets. *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, 24-28 June 2007
D0123	A. Kumar, R. Kumar, P.S. Grover. Towards a Unified Framework for Cohesion Measurement in Aspect-Oriented Systems. *Proceedings of the 19th Australian Conference on Software Engineering*. ASWEC 2008. pp.57-65.	O0142	Thiago T. Bartolomei, Alessandro Garcia, Claudio Sant'Anna & Eduardo Figueiredo. Towards a Unified Coupling Framework for Measuring AspectOriented. *Proceedings of the 3rd International Workshop on Software Quality Assurance* (Portland,Oregon, November 06 - 06, 2006). SOQUA '06.
D0124	Reza Khajavinia. Managing Process Interaction in IS Development Projects. *Proceedings of the Fifth International Conference on Information Technology: New Generations*, April 2008.	O0143	Bendik Bygstad & Peter Axel Nielsen. Understanding and Managing Process Interaction in IS Development Projects. *Proceedings of the 28th Information Systems Research Seminar in Scandinavia* (IRIS), Kristiansand, August 2005
D0125	A. Dana, A.M.Yadegari, A. Salahi, S. Faramehr, H. Khosravi. A New Scheme for on-Demand Group Mobility Clustering in Mobile Ad hoc Networks. *Proceedings of the 10th International Conference on Advanced Communication Technology*, 2008. (ICACT '08), Vol 2, 2008, pp. 1370-	O0144	Curt Cramer, Oliver Stanze, Kilian Weniger & Martina Zitterbart. Demand-Driven Clustering in MANETs. *Proceedings of the International Conference on Wireless Networks*, 2004. (ICWN '04), Vol. 1, 2004, pp. 81-87
D0126	Vinoth Sivasubramaniam. On the Safety and Efficiency of Firewall Policy Deployment. *Proceedings of the IEEE Network Operations and Management Symposium Workshops*, 2008, pp.138-152.	O0145	Charles C. Zhang, Marianne Winslett & Carl A. Gunter. On the Safety and Efficiency of Firewall Policy Deployment. *Proceedings of the 2007 IEEE Symposium on Security and Privacy* (SP07), May 2007, pp.33-50

D0127	T. Hammons and J. Schwarz. Europe: Recent Developments of Transmission System Interconnections and the Implementation of Power Generation with Respect to the Kyoto Protocol. *Proceedings of 2005 IEEE Russia Power Tech*, 27-30 June 2005	O0146	W. Leonhard & M. Grobe . Sustainable Electrical Energy Supply with Wind and Pumped Storage - a Realistic Long-term Strategy or Utopia?. *Proceedings of the 2004 IEEE Power Engineering Society General Meeting*, June 2004.
		O0147	Jiri Feist. Status of Resynchronization of the two UCTE Synchronous zones. *Proceedings of the 2004 IEEE Power Engineering Society General Meeting*, June 2004.
		O0148	Bruno Cova. Progress of the Mediterranean Ring and the interconnection with Europe. *Proceedings of the 2004 IEEE Power Engineering Society General Meeting*, June 2004.
		O0149	Y. Kucherov, L. Koshcheev, and Y. Tikhonov. Technical Rules and Standards Harmonization is a Key Issue for Realization of UCTE, NORDEL and IPS/UPS Synchronous Operation. *Proceedings of the 2004 IEEE Power Engineering Society General Meeting*, June 2004.
D0128	Ahmed Zahoor, J.P. Cances, and V. Meghdadi. Beamforming Applied to New Cooperative Decode-and-Forward Relay Schemes. *Proceedings of the IEEE Vehicular Technology Conference*, Spring 2008	O0150	Jian Zhao, Marc Kuhn, Armin Wittneben & Gerhard Bauch. Cooperative Transmission Schemes for Decode-and-Forward Relaying. *Proceedings of the 18th Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2007, pp 1-5
D0129	Min Gan, Hongfa Wang, Guoying Yue, Nan Xie, Haibo Zhang. Observations on Liner Prediction Based Blind Source Extraction Algorithms. *Proceedings of the Seventh IEEE/ACIS International Conference on Computer and Information Science* (ICIS 2008), May 2008, pp. 373-376	O0151	Zhi-Lin Zhang & Liqing Zhang. Linear Prediction Based Blind Source Extraction Algorithms in Practical Applications. M.E. Davies et al. (Eds.). *Proceedings of the 7th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2007),* 2007, pp. 309-316, pp. 309–316.
D0130	A. Bazregar, A. Movaghar, A. Barati, M.R.E. Nejhad, H. Barat. A New Automatic Clustering Algorithm via Deadline Timer for Wireless Ad-hoc Sensor Networks. *Proceedings of the 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, 2008. ICTTA 2008. pp.1-6	O0152	Chih-YuWen & WilliamA. Sethares. Automatic Decentralized Clustering forWireless Sensor Networks. *EURASIP Journal onWireless Communications and Networking*, 2005:5, pp. 686–697.
D0131	Po-Yi Shih; Jhing-Fa Wang; Hsiao-Ping Lee; Hung-Jen Kai; Hung-Tzu Kao; Yuan-Ning Lin. Acoustic and Phoneme Modeling Based on Confusion Matrix for Ubiquitous Mixed-Language Speech Recognition. *Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing*, 2008.	O0153	Matti Vihola. *Dissimilarity Measures for Hidden Markov Models and Their Application in Multilingual Speech Recognition*. Master of Science Thesis, Tampere University of Technology, 2001

	SUTC '08, pp. 500-506.		
D0132	Yanping Gao, Changhui Deng, Yandong Che. An Adaptive Index-based Algorithm using Time-coordination in Mobile Computing. *Proceedings of the 2008 International Symposium on Information Processing (ISIP 08*), May 2008, pp. 578-585.	O0154	Awadhesh Kumar Singh. On Mobile Checkpointing using Index and Time Together. *Proceedings of World Academy of Science, Engineering and Technology*, Vol 26, December 2007, pp.144-151.
D0133	Youssef A. Mobarak. Modified Load Flow Analysis for Integrated AC/DC Power Systems. *12th International Middle-East Power System Conference (MEPCON 2008)*, 2008, pp. 402–405	O0155	A. Panosyan & B. R. Oswald. Modified Newton-Raphson Load Flow Analysis For Integrated AC/DC Power Systems. *Proceedings of the 39th International Universities Power Engineering Conference (UPEC 2004)*, 2004, pp. 1223-1227.
D0134	Yong-Sung Choi, Ju-Ho Yun & Kyung-Sup Lee. On-line Insulation Diagnostic System and Off-line PD Monitoring with HVAC Testing. *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008, pp 85-88.	O0156	B. Kübler & W. Hauschild. What to do after the Warning of an On-line Insulation Diagnostic System? Off-line PD Monitoring with HVAC Testing!. *Proceedings of the 2004 Power Systems Conference and Exposition, 10-13 Oct.2004*, vol. 2, pp. 822-828.
D0135	Yong-Sung Choi, Ju-Ho Yun, Hyang-Kon Kim, Chung-Seog Choi & Kyung-Sup Lee. Charge Transport and Electroluminescence in Insulating Polymers. *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008, pp. 238-242.	O0157	C. Laurent, G. Teyssedre & S . Le Roy. A discussion on charge transport and electroluminescence in insulating polymers. *Proceedings of the 2005 International Symposium on Electrical Insulating Materials*, 5-9 June 2005, vol. 1, pp. 7-15.
D0136	Yong-Sung Choi, Ju-Ho Yun, Hyang-Kon Kim, Chung-Seog Choi & Kyung-Sup Lee. On-line Condition Monitoring and Diagnostics of Power Distribution Equipment. *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008, pp 692-695.	O0158	Albert Livshitz, Bella H. Chudnovsky & Boris Bukengolts. On-Line Condition Monitoring and Diagnostics of Power Distribution Equipment. *Proceedings of the 2004 Power Systems Conference and Exposition*, 10-13 Oct. 2004, vol. 2, pp. 646-653.
D0137	Yong-Sung Choi, Ju-Ho Yun & Kyung-Sup Lee. Wireless Sensor for Managing Electrical Distribution Networks. *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008, pp 825-829.	O0159	Mikael M. Nordman & Matti Lehtonen. A Wireless Sensor Concept for Managing Electrical Distribution Networks. *Proceedings of the 2004 Power Systems Conference and Exposition*, 10-13 Oct. 2004, vol. 2, pp. 1198-1206.
D0138	Yong-Sung Choi, Ju-Ho Yun & Kyung-Sup Lee. Automatic Vacuum Capacitor Switch with Modified Digital Filter. *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008, pp. 846-849.	O0160	P.S.Harish & G.Prashant. Automatic Vacuum Capacitor Switch With Modified Digital Filter Design For Enhanced Speed And Power Optimization. *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008, pp. 846-849.

D0139	Yong-Sung Choi, Ju-Ho Yun & Kyung-Sup Lee. Practical Implications of Multi-agent Transformer Condition Monitoring. *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008, pp 945 - 948	O0161	V. M. Catterson & S. D. J. McArthur. The Practical Implications of Bringing a Multi-Agent Transformer Condition Monitoring System On-Line. *Proceedings of the 2004 Power Systems Conference and Exposition*, 10-13 Oct. 2004, vol. 1, pp.12 - 16.
D0140	Zheng Chen, Xiaojing Wang, Yili Jin, Honglei Zhou. Exploring Fault-tolerant Distributed Storage System using GE code. *Proceedings of the 2008 International Conference on Embedded Software and Systems*, July 2008, pp. 142-148.	O0162	Benjamin Gaidioz, Birger Koblitz & Nuno Santos. Exploring high performance distributed file storage using LDPC codes. *Parallel Computing*, 33 (2007), pp. 264–274.
D0141	K. Kaur, S. Chowdhury, S.P. Chowdhury, K.B. Mohanty, A. Domijan. Fuzzy Logic Based Control of Variable Speed Induction Machine Wind Generation System. *Proceedings of the IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, 2008 pp.1-11.	O0163 O0164	M. Godoy Simoes & Bimal K. Bose. Fuzzy Logic Based Intelligent Control of a Variable Speed Cage Machine Wind Generation System. *EEE Transactions on Power Electronics*, vol.12, no.1, Jan 1997, pp. 87-95. M. Godoy Simoes, Bimal K. Bose & Ronald J. Spiegel. Design and Performance Evaluation of a Fuzzy-logic-based Variable-speed Wind Generation System. *IEEE Transactions on Industry Applications*, vol.33, no.4, Jul/Aug 1997, pp. 956-965,.
D0142	Y.Shi and A. Monti. FPGA-based Fast Real Time Simulation of Power Systems. *Proceedings of the 2008 Power and Energy General Meeting*, July 2008.	O0165	Mahmoud Matar & Reza Iravani. An FPGA-based Real-Time Digital Simulator for Power Electronic Systems. *International Conference on Power Systems Transients (IPST'07)* in Lyon, France, June 4-7, 2007
D0143	ZAHOOR AHMED, J. P. CANCES AND V. MEGHDADI. On Capacity of Gaussian 'Light' Relay Channel. *Proceedings of the International Wireless Communications and Mobile Computing Conference*, 2008 (IWCMC '08) 6-8 Aug. 2008	O0166	Mohammad Ali Khojastepour, Ashutosh Sabharwal & Behnaam Aazhang. On Capacity of Gaussian 'Cheap' Relay Channel. *Proceedings of the IEEE Global Telecommunications Conference*, 2003 (GLOBECOM'03) Vol. 3, 1-5 Dec. 2003, pp. 1776-1780.
D0144	Zhang Hu. Algorithm for the GME Problem in Real-Time Distributed Systems Based on Token. *Proceedings of the 3rd International Conference on Innovative Computing Information and Control, 2008* (ICICIC '08), pp. 206-206,	O0167	Abhishek Swaroop & Awadhesh Kumar Singh. A Distributed Group Mutual Exclusion Algorithm for Soft Real Time Systems. *World Academy of Science, Engineering and Technology*, vol 26, December 2007, pp. 138-143.
D0145	Jian-hui Wang and Chun-zhang Bai. Software Evolution with Feature-Oriented and Aspect-Oriented Programming. *3rd International Conference on Innovative Computing Information and Control*, 2008. ICICIC 2008.	O0168	Sven Apel, Thomas Leich, Marko Rosenmuller & Gunter Saake. Combining FeatureOriented and Aspect Oriented Programming to Support Software Evolution. *Proceedings of the 2nd ECOOP Workshop on Reflection, AOP and Meta-Data for Software Evolution* (RAM-SE), pp 3–16.

D0146	Jiemin Liu, Hongxing Zou, Jingxin Dou, and Yuan Gao. Asymmetric Path Optimization In Mobile Multi-homed SCTP Multimedia Transport. *Proceedings of the International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (IIH-MSP/2008) pp 496-499	O0169	Eduardo Parente Ribeiro & Victor C. M. Leung. Asymmetric Path Delay Optimization in Mobile Multi-homed SCTP Multimedia Transport. *WMuNeP'05*, October 13, 2005, Montreal, Quebec, Canada.
D0147	Jian-hui Wang, Yan Jiang, and Hai-long Zhang. An Implementation of Fuzzy Clustering with Size and Shape Constraints. *Proceedings of the 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp.922-925	O0170	Christian Borgelt & Rudolf Kruse. Fuzzy and Probabilistic Clustering with Shape and Size Constraints. *Proceedings of the 11th International Fuzzy Systems Association World Congress (IFSA'05*, Beijing, China), 945-950.
D0148	Tian Feng-min, Xu Ding-jie, Zhao Yu-xin and Li Ning. Underwater Vehicle Terrain Navigation Based on Maximum Likelihood Estimation. *Proceedings of the 2008 International Conference on Information and Automation*, pp. 1268-1273.	O0171	Ingemar Nygren. *Terrain Navigation for Underwater Vehicles*. Doctoral Thesis in Signal Processing, Royal Institute of Technology, Stockholm, Sweden 2005.
D0149	Shujuan Peng, Yuanxiang Li & Hui Zeng. Efficient Collision Detection for Arbitrary Polyhedral Model based on GPU and Swept Volume. *Proceedings of the International Colloquium on Computing, Communication, Control and Management*, ISECS, Vol 2, August 2008, pp. 260-264.	O0172	Zhaowei Fan,HuagenWan & Shuming Gao. IBCD: a fast collision detection algorithm based on image space usingOBB. *Journal of Visualization and Computer Animation* (14) John Wiley and Sons, September 2003, pp. 169-181.
D0150	A.J.M Adnan, R.Mohamad, I.A.Tengku, S.Shaari. 1310/1550 nm Photonic Crystal Based on Multimode Interference Demultiplexer. *Joint Conference of the Opto-Electronics and Communications Conference*, 2008 and the 2008 Australian Conference on Optical Fibre Technology. OECC/ACOFT 2008.	O0173	Hyun-Jun Kim, Insu Park, Beom-Hoan O, Se-Geun Park, El-Hang Lee & Seung-Gol Lee. Self-imaging phenomena in multi-mode photonic crystal line-defect waveguides: application to wavelength de-multiplexing. *Optics Express*, Vol. 12, No 23, Nov 2004.
D0151	Hong-Zhen Zheng, Dian-Hui Chu, De-Chen Zhan and Xiao-Fei Xu. Dual Sales Channel Management with Service Competition. *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*, July 2008	O0174	Kay-Yut Chen, Murat Kaya & Özalp Özer. Dual Sales Channel Management with Service Competition. *Manufacturing & Service Operations Management*, Vol 10, No 4, April 2008, pp 654-675.
D0152	Dehuai Zeng, Gang Xu, Cunxi Xie, Degui Yu. Artificial Immune Algorithm based Robot Obstacle-Avoiding Path Planning. *Proceedings of the IEEE International Conference on Automation and Logistic*s, September 2008, pp. 798-803.	O0175	Guan-Chun Luh & Wei-Wen Liu. An Immunological Approach to Mobile Robot Navigation. *Applied Soft Computing*, Vol 8, No 1, December 2006, Elsevier, pp. 30-45
D0153	Omid Jadidi, Tang Sai Hong, Fatemeh Firouzi, and Rosnah Binti Mohd Yusuff. A Grey Based Method Based on TOPSIS Concepts for Multiple Criteria Decision Making Problems. *4th IEEE International Conference on Management of Innovation and Technology*, 2008.	O0176	Guo-Dong Lia, Daisuke Yamaguchia & Masatake Nagai. A grey-based decision-making approach to the supplier selection problem. *Mathematical and Computer Modelling* 46 (2007) 573–581

D0154	Chengzhu Sun, Xiaofei Xu, Xiangyang Li, Shengchun Deng. Knowledge Discovery from Virtual Enterprise Model Based on Semantic Annotation. *Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, 2008, vol.5, pp.546-551.	O0177	M.Missikoff, F.Schiappelli & F.Taglino. A Controlled Language for Semantic Annotation and Interoperability in e-Business Applications. *Proceedings of the Semantic Integration Workshop, collocated with the Second International Semantic Web Conference (ISWC-03)*, October 2003
D0155	Mochammad Zuliansyah, Suhono Harso Supangkat, Yoga Priyana, Carmadi Machbub. 3D Topological Relations for 3D Spatial Analysis. *Proceedings of the 2008 IEEE Conference on Cybernetics and Intelligent Systems*, pp. 585-590	O0178	Siyka Zlatanova. On 3D Topological Relationships. *11th International Workshop on Database and Expert System Applications*, DEXA 2000, pp. 913-919.
D0156	Hong Shi. A New Private Multi-Channel UWB MAC Protocol Design for Distributed Network. *Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing*, 2008.	O0179	Zhi-hong Xiao, Yao-hui Wu & Zong-qi Guan. A New Private Multi-Channel UWB MAC Protocol Design for Distributed Network. *Proceedings of the First International Workshop on Knowledge Discovery and Data Mining, 2008*. WKDD 2008. Jan. 2008, pp. 489-492.
D0157	Zhaojian Liu, Guangqi Sun, Qing Li. Implications for Transport Systems by Emerging Global Logistics Networks. *4th International Conference on Wireless Communications, Networking and Mobile Computing 2008* (WiCOM 08), 2008, pp. 1–4.	O0180	Lorant Tavasszy, C.J. Ruijgrok, M.J.P.M. Thissen. Emerging Global Logistics Networks: Implications for Transport Systems and Policies. *Growth and Change*, Vol. 34 No. 4 (Fall 2003), pp. 456-472.
D0158	Shengbo Hu, Xin Meng. Classification and Key Management Approaches for Space Networks Security. *Proceedings of the 2008 International Conference on Anti-Counterfeiting, Security and Identification*, August 2008	O0181	Aruna Balasubramanian & Sumita Mishra. Secure Key Management for NASA Space Communication. *Proceedings of the 2005 ICNS Conference and Workshop*, May 2005.
D0159	Xianming Chen, Zhengxun Song, Zhen Hu, Bin Guo. An Improved CABS Capture Algorithm for PAM-UWB. *Proceedings of the IEEE 9th International Conference on Signal Processing, 2008 (ICSP 2008*), pp. 1955-1958.	O0182	Yeqiu Ying, Mounir Ghogho & Ananthram Swam. Analysis of Code-Assisted Blind Synchronization for UWB Systems. *Proceedings of the IEEE International Conference on Communications*, 2007. (ICC 2007), June 2007, pp. 4293-4298
D0160	A. Kumar, R. Kumar, P.S. Grover. Towards a Unified Framework for Complexity Measurement in Aspect-Oriented Systems. *Proceedings of the 2008 International Conference on Computer Science and Software Engineering*, vol.2, pp.98-103.	O0142	Thiago T. Bartolomei, Alessandro Garcia, Claudio Sant'Anna & Eduardo Figueiredo. Towards a Unified Coupling Framework for Measuring AspectOriented. *Proceedings of the 3rd International Workshop on Software Quality Assurance* (Portland,Oregon, November 06 - 06, 2006). SOQUA '06.
D0161	Shenjun Xue, Yi Liu. A CSCW Systems Classification. *Proceedings of the 2008 International Conference on Computer Science and Software Engineering*, December 2008, pp. 996-999	O0184	V.M.R. Penichet, I. Marin, J.A. Gallud, M.D. Lozano & R. Tesoriero. A Classification Method for CSCW Systems. *Proceedings of the Second International Workshop on Views on Designing Complex Architectures*, February 2007, pp. 756-773 / Electronic Notes in Theoretical Computer Science 168 (2007) 237–247

D0162	Gang Yi. Research on a Novel MAS for Automated Negotiation Method. *Proceedings of the 2008 3rd International Conference on Intelligent System and Knowledge Engineering*, November 2008, pp. 635-639	O0056	Weijin Jiang & Xuancheng Zhou. Research on a Novel Multi-Agent System Negotiation Strategy and Model. *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing. WiCOM '08*, October 2008.
D0163	Dejia Shi, Weijin Jiang, Danping Xu. Research on a Novel Multicast Routing Hybrid Optimization Algorithm and Its Application. *Proceedings of the 2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing (EUC 2008)* December 2008, pp. 339-345	O0186	Chaoliang Li & Liang Wei. Research on QoS Multicast Routing Optimization Algorithm Based on Hybrid Genetic Algorithm. *Proceedings of the 2008 4th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM 2008).*
D0164	M. Sayfullah, B. Roland, A.L. Scholtz. Jitter Analysis of a Mixed PLL-DLL Architecture. *Proceedings of the International Conference on Electrical and Computer Engineering, 2008. ICECE* 2008, pp. 750-754.	O0187	Han-Yuan Tan. *Design of Noise Robust and Data Recovery Using an Adaptative Bandqidth Mixed PLL/DLL.*. Harvard University, 2006
D0165	Jian Wen Tao and Pei Fen Ding. CTMIR: A Novel Correlated Topic Model for Image Retrieval. *Proceedings of the Second International Workshop on Knowledge Discovery and Data Mining*, WKDD 2009 pp.948-951.	O0188	T. Greif, E. Horster & R. Lienhart. *Correlated Topic Models for Image Retrieval*. Institut fur Informatik, Universitat Augsburg, Report 2008-09
D0166	Dong Li, Guiyou Chen. A Wide Bandwidth Current Probe Based on Rogowski Coil and Hall Sensor. *Proceedings of the 5th International CES/IEEE Power Electronics and Motion Control Conference 2006 (*IPEMC 2006), p. 1-5	O0189	Nicolas Karrer, Patrick Hofer-Noser & Daniel Henrard. A New Current Probe with a Wide Bandwidth. *Proceedings of the European Conference on Power Electronics and Applications 1999 (EPE 99)*, Lausanne, Switzerland.
D0167	Chenxia Sun, Xiaoyang He, Chao Wang. The Web-based B2B Environment with Web Services. *Proceedings of the 2009 International Conference on Electronic Computer Technology*, February 2009, pp. 103-107.	O0190	S. Karetsos, C. Costopoulos, O. Pyrovolakis & L. Georgiou. Developing Agricultural B2B Processes Using Web Services. *Proceedings of the 6th WSEAS Int. Conf. on Software Engineering, Parallel and Distributed Systems*, Corfu Island, Greece, February 16-19, 2007, pp. 161-168.
D0168	S.F. Islam, A.H.M.S. Islam, M.M.R. Billah. Active Noise Control for Industrial Applications. *Proceedings of the 11th International Conference on Computer and Information Technology*, December 2008, pp. 312-316.	O0191	Benny Sallberg, Lars Hakansson & Ingvar Claesson. Active Noise Control for Hearing Protection using a Low Power Fixed Point Digital Signal Processor. *Proceedings of the International Workshop on Acoustic Echo and Noise Control*, September 2005, pp. 65-68.
D0169	Md. Saidur Rahman, G. M. Atiqur Rahaman, Asif Ahmed and G. M. Salahuddin. An Approach to Recognize Handwritten Bengali Numerals for Postal Automation. *Proceedings of 11th International Conference on Computer and Information Technology*, (ICCIT 2008, 25-27 December, 2008, Khulna, Bangladesh)	O0192	YingWena,Yue Lub & Pengfei Shi. Handwritten Bangla numeral recognition system and its application to postal automation. *Pattern Recognition* 40 (2007), pp. 99 – 107

D0170	H.M. Khodr, Zita A. Vale, Carlos Ramos. A Benders Decomposition and Fuzzy Multicriteria Approach for Distribution Networks Renumerations Considering DG. *IEEE Transactions on Power Systems*, Vol 24, No 2, May 2009.	O0193	P. M. De Oliveira-De Jesús, M. T. Ponce de Leão & H.M Khodr. Remuneration of Distribution Networks using a Fuzzy Multicriteria Planning Algorithm. *9th International Conference on Probabilistic Methods Applied to Power Systems KTH*, Stockholm, Sweden – June 11-15, 2006
D0171	Zhi Teng, Ye Liu, Fuji Ren. A Foundation for Knowledge System with Application in Information Retrieval and Knowledge Acquisition. *Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering*, 2008. NLP-KE '08.	O0194	Lucy Vanderwendea, Hisami Suzukia, Chris Brocketta & Ani Nenkova. Beyond SumBasic: Task-Focused Summarization with Sentence Simplification and Lexical Expansion. *Information Processing and Management*, Vol 43, Elsevier, 2007, pp 1606-1618.
		O0195	Lawrence H. Reeve, Hyoil Han & Ari D. Brooks. The use of domain-specific concepts in biomedical text summarization. *Information Processing and Management*, 43 (2007), pp. 1765–1776
D0172	Zhou Xusheng and Wang Zhiming. Application of Markov Chain in IP Traffic Classification. *Proceedings of the International Conference on Networks Security, Wireless Communications and Trusted Computing*, 2009. NSWCTC '09. , vol.2, pp.688-691.	O0196	Hamza Dahmouni, Sandrine Vaton & David Rossé. A Markovian Signature-Based Approach to IP Traffic Classification. *Proceedings of the 3rd Annual ACM Workshop on Mining Network Data*, MineNet '07, pp. 29-34.
D0173	Tang Shu-Pian. Resonant Energy Transfer from Organics to Quantum Dots and Carrier Multiplication. *Proceedings of the 2009 Asia-Pacific Power and Energy Engineering Conference*, March 2009.	O0197	Vladimir M. Agranovich & Gerard Czajkowski. *Resonant Energy Transfer from Organics to Quantum Dots and Carrier Multiplication*. arXiv:0801.3794v1, 24 January 2008
D0174	Nasrullah Memon, Abdul Rasool Qureshi, Uffe Kock Wiil, and David L. Hicks. Novel Algorithms for Subgroup Detection in Terrorist Networks. *Proceedings of the 2009 International Conference on Availability, Reliability and Security*, March 2009, pp. 572-577.	O0198	Adam Perer & Ben Shneiderman. Balancing Systematic and Flexible Exploration of Social Networks. *EEE Transactions on Visualization and Computer Graphics*, Vol. 12, No. 5, Sept/Oct 2006, pp. 693-700
		O0199	Fatih Ozgu, Julian Bondy & Hakan Aksoy. Mining for offender group detection and story of a police operation. *Proceedings of the Sixth Australasian Data Mining Conference* (AusDM 2007), Gold Coast, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 70, December 2007, pp. 185-189.
		O0200	M. N. Smith & P. J. H. King. The Exploratory Construction of Database Views. *Research Report BBKCS-02-02*, School of Computer Science and Information Systems, Birbeck College, University of London. 2002.

D0175	Xinyu Zhang, Wei Tang, Cheng Sun. The Web-based B2B Environment with Web Services. *Proceedings of the 2009 Sixth International Conference on Information Technology: New Generations*, June 2009, pp. 1623-1624.	O0190	S. Karetsos, C. Costopoulos, O. Pyrovolakis & L. Georgiou. Developing Agricultural B2B Processes Using Web Services. *Proceedings of the 6th WSEAS Int. Conf. on Software Engineering, Parallel and Distributed Systems*, Corfu Island, Greece, February 16-19, 2007, pp. 161-168.
D0176	G.K. Santhalia, N. Sharma, S. Singh, M. Das, J. MulChandani. A Method to Extract Future Warships in Complex Sea-Sky Background which May Be Virtually Invisible. *Proceedings of the Third Asia International Conference on Modelling & Simulation*, May 2009.	O0201	W J Kang, X M Ding, J W Cui & L Ao. Research on Extraction of Ship Target in Complex Sea-sky Background. *International Symposium on Instrumentation Science and Technology, Journal of Physics: Conference Series 48*, 2006, pp. 354-358.
D0177	Baoping Xiao, Chang Xu, and Lijun Xu. System Model and Controller Design of an Inverted Pendulum. *Proceedings of the 2009 Second International Conference on Industrial and Information System*s, April 2009, pp 356-359.	O0202	John Stang. *The Inverted Pendulum*. Design Project Report, Cornell University, May 2005..
D0178	Baoping Xiao, Chang Xu and Lijun Xu. An Application of MEMS Sensors in Inertial Navigation Systems. *Proceedings of the 2009 Second International Conference on Industrial and Information Systems*, April 2009, pp 360-363.	O0203	Jordan Crittenden & Parker Evans. *MEMS Inertial Navigation System*. Student Project: ECE4910, Cornell University, May 8, 2008.
D0179	Yuhua Zhu. A Novel View Multi-view Synthesis Approach for Free Viewpoint Video. *Proceedings of the 2009 International Joint Conference on Artificial Intelligence* April 2009, pp. 88-91.	O0204	Eddie Cooke, Peter Kauff & Thomas Sikor. Multi-view synthesis: A novel view creation approach for free viewpoint video. *Signal Processing: Image Communication*, 21 (2006), pp. 476–492.
D0180	Tong Zhen and Qiuwen Zhang. A Combining Heuristic Algorithm for the Multi-depot Vehicle Routing Problem with Inter-Depot Routes. *Proceedings of the 2009 International Joint Conference on Artificial Intelligence*, April 2009, pp. 436-439.	O0205	Benoit Crevier, Jean-François Cordeau & Gilbert Laporte. The Multi-Depot Vehicle Routing Problem with Inter-Depot Routes. *European Journal of Operational Research*, Vol. 176, Issue 2, January 2007, pp. 756-773.
D0181	Napur Lodha, Nivesh Rai, Aarthy Krishnamurthy, Hrishikesh Venkataraman. Efficient Implementation of QRD-RLS Algorithm using Hardware-Software Co-design. *2009 IEEE International Symposium on Parallel & Distributed Processing* (IPDPS 2009), 2009, pp. 1 – 4.	O0206	Nupur Lodha, Nivesh Rai, Rahul Dubey & Hrishikesh Venkataraman. Hardware-Software Co-design of QRD-RLS Algorithm with Microblaze Soft Core Processor. S.K. Prasad et al. (Eds.): *ICISTM 2009*, CCIS 31, pp. 197–207.
D0182	D. Hussain, S. Ather. Reproducible Research in Various Facets of Signal Processing. *Proceedings of the International Multimedia, Signal Processing and Communication Technologies*, 2009. IMPACT '09, pp. 64 – 68	O0207	P. Vandewalle, G. Barrenetxea, I. Jovanovic, A. Ridolfi & M. Vetterli. Experiences with Reproducible Research in Various Facets of Signal Processing Research. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*. Volume: 4, page(s): IV-1253 - IV-1256

D0183	S. Paramasivam, and M. Kumaran. Evaluation of GP Model for Software Reliability. *2009 International Conference on Signal Processing Systems*, May 2009	O0208	Wasif Afzal & Richard Torkar. A comparative evaluation of using genetic programming for predicting fault count data. *Third International Conference on Software Engineering Advances*, 2008. ICSEA '08, pp.407-414.
D0184	Hou Ming. Process Control Technology Application in the Manufacture of Discrete Component NC. *Proceedings of 2009 World Congress on Computer Science and Information Engineering*.	O0209	Sanjeev Kumara, Aydin Nassehia, Stephen T. Newman, Richard D. Allen, Manoj K. Tiwari. Process control in CNC manufacturing for discrete components A STEP-NC compliant framework. *Robotics and Computer-Integrated Manufacturing,* Vol 23, No 6, December 2007, Elsevier, pp. 667-676.
D0185	Kanak Saxena. Efficient Mining of Weighted Temporal Association Rules. *Proceedings of the 2009 World Congress on Computer Science and Information Engineering*, March 2009, pp. 421-425	O0210	Chang-Hung Lee. *Mining Association Relationship in a Temporal Database*. PhD Thesis, National Taiwan University, Taipei, Taiwan, May 2002.
		O0211	Mireille Samia. A Representation of Time Series for Temporal Web Mining. *Proceedings of the 16th GI-Workshop on the Foundation of Databases*, June 2004, pp. 103-107
D0186	Zhikuan Sun & Shiwei Lin. Study on Application of Fuzzy Theory for Knowledge Management in E-Commerce. *Proceedings of the International Symposium on Information Engineering and Electronic Commerce*, 2009, IEEC '09, pp.450-453, May 2009	O0212	Wei long Liu, Xin Zhang & Fang Jin. Fuzzy Knowledge for Agent Oriented Knowledge Management in E-Commerce. *Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, 2008, FSKD '08, vol.5, pp.175-179.
D0187	Xuefeng Cao, Gang Wan & Feng Li. 3D Vector-Raster Data Integration Model Based on View Dependent Quadtree and GPU Friendly Rendering Algorithm. *Proceedings of the International Joint Conference on Computational Sciences and Optimization, CSO 2009*, Vol.2, pp.244-247.	O0213	Eric Bruneton & Fabrice Neyret. Real-time rendering and editing of vector-based terrains. *Computer Graphics Forum 27*, 2 (2008), pp. 311-320.
D0188	Tong Zhen, Yuhua Zhu & Qiuwen Zhang. A RFID Logistics Resource Management System for the Warehouses. *Proceedings of the 2009 International Conference on Environmental Science and Information Application Technology*, July 2009, pp. 63-66	O0214	T.C. Poon, K.L. Choy, Harry K.H. Chow, Henry C.W. Lau, Felix T.S. Chan & K.C. Ho. A RFID case-based logistics resource management system for managing order-picking operations in warehouses. *Expert Systems with Applications*, 36, May 2009, pp 8277–8301.
D0189	Shao-Gang Dong, Zhong-Hua Tang, Bai-Wei Liu & O.D. Orodu. Numerical Modelling of the Environment Impact of Landfill Leachate Leakage on Groundwater Quality--A Field Application. *Proceeedings of the 2009 International Conference on Environmental Science and Information Application Technology*, ESIAT 09, pp. 565-568	O0215	M. P. Papadopoulou, G. P. Karatzas & G. G. Bougioukou. Numerical modelling of the environmental impact of landfill leachate leakage on groundwater quality – a field application. *Environmental Modeling and Assessment*, 12:1, February 2007, pp 43-54.

D0190	Dejia Shi & Weijin Jiang. Robust Guaranteed Cost Control Base on T-S Fuzzy Model. *Proceedings of the 2009 International Conference on Measuring Technology and Mechatronics Automation(ICMTMA 2009*), August 2009, pp. 586-589	O0216	Rui Yang, Tianmin Huang, Cuihong Wang & Li Zou. Fuzzy Guaranteed Cost Control for Uncertain Singular Systems with State and Input delays. *Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering (ISKE 2007)*, Atlantis Press, October 2007.
D0191	M. Sultan, M. Siddiqui, A.H. Sajid, D.G. Chougule. FPGA Based Efficient Implementation of PID Control Algorithm. *Proceedings of the International Conference on Control, Automation, Communication and Energy Conservation*, June 2009	O0217	Yuen Fong Chan, M. Moallem & Wei Wang. Design and Implementation of Modular FPGA-based PID Controllers. *IEEE Transactions on Industrial Electronics*, vol.54, no.4, Aug. 2007, pp. 1898–1906.
D0192	Luan Jia-hui, Tao Lai Fa & Lu Chen. Sensor Fault Reconstruction for Satellite Gyroscope Using Eigenstructure Assignment. *Proceedings of the WRI Global Congress on Intelligent Systems, 2009 (GCIS '09)*, vol.2, May 2009, pp. 318-322	O0218	Chee Pin Tant & Maki K. Hahibt. Robust sensor fault reconstruction for an inverted pendulum using right eigenstructure assignment. *Proceedings of the 2004 IEEE Intemational Conference on Control Applications* Taipei. Taiwan, September 2004, pp. 1236- 1241.
D0193	Yongjian Tao & Decun Dong. An Improved Modular Approach for System Reliability Analysis. *Proceedings of the 2009 WASE International Conference on Information Engineering*, November 2009, pp. 111-113.	O0219	H. Boudali & J.B. Dugan. A discrete-time Bayesian network reliability modeling and analysis framework. *Reliability Engineering and System Safety*, 87, March 2005, pp. 337–349.
D0194	Fernando Carvalho, Silvio R.L. Meira, Bruno Freitas & Nathalia Carvalho. An Embedded Software Component Quality Model—EQM. *Proceedings of the IEEE International Conference on Information Reuse & Integration*, 2009, IRI '09., pp. 444-445.	O0220	Alexandre Alvaro, Eduardo Santana de Almeida & Silvio Lemos Meira. A Software Component Quality Model: A Preliminary Evaluation. *Proceedings of the 32nd EUROMICRO Conference on Software Engineering and Advanced Applications (EUROMICRO-SEAA'06)*
D0195	Jian-Min Zhang & Lei Li. A Support Vector Machine Approach for Edge Detection in Noisy Images. *Proceedings of the Eighth International Conference on Machine Learning and Cybernetics*, 12-15 July 2009	O0221	Hilario Gómez-Moreno, Saturnino Maldonado-Bascón & Francisco López-Ferreras. Edge Detection in Noisy Images Using the Support Vector Machines. *Lecture Notes in Computer Science*. Vol. 2084, pp. 685-692, 2001.
		O0222	H. Gómez Moreno, S. Maldonado Bascón, F. López Ferreras, P. Gil Jiménez. A New and Improved Edge Detector Using the Support Vector Machines. *Advances in Systems Engineering, Signal Processing and Communications*, 2002, pp. 239-243.
D0196	Xu Yaun & Tang Shu-Pian. Supersolid Behavior of Nonlinear Light. *Publication Symposium on Photonics and Optoelectronics*, 2009 (SOPO 2009), pp. 1-4.	O0223	Albert Ferrando, Miguel-Angel Garcia-March & Mario Zacaroes. *Supersolid behavior of nonlinear light*. arXiv : 0808.0998v1, 7 Aug 2008.

D0197	Yongjian Tao, Decun Dong, Peng Ren. Decision Trees Generation Based on Fault Trees Analysis. *Proceedings of the 2009 International Forum on Information Technology and Applications*, 2009, August 2009, pp. 178-180.	O0224	Zhihua Tang & Joanne Bechta Dugan. Minimal Cut Set/Sequence Generation for Dynamic Fault Trees. *Proceedings of the 2004 Annual Symposium Reliability and Maintainability*, Jan. 2004 pp. 207- 213.
		O0225	Tariq Assaf & Joanne Bechta Dugan. Diagnostic Expert Systems from Dynamic Fault Trees. *Proceedings of the 2004 Annual Symposium Reliability and Maintainability*, Jan. 2004, pp. 444-450.
D0198	Xuefeng Cao, Gang Wan, and Feng Li. 3D Real-Time Representation of Sociocultural Features Based on Vector-Raster Data Integration Model. *Proceedings of the International Forum on Information Technology and Applications*, 2009. IFITA '09. Vol.3, pp.587-590.	O0213	Eric Bruneton & Fabrice Neyret. Real-time rendering and editing of vector-based terrains. *Computer Graphics Forum 27*, 2 (2008), pp. 311-320.
D0199	WenJie Tian, Yu Geng, Lan Ai, JiCheng Lu. Dynamic Compensation for Infrared Thermometer Based on Wiener Model and Immune Clone Selection Algorithm. *2009 Pacific-Asia Conference on Circuits, Communications and Systems (PACCS 2009)*, pp 301 – 305.	O0226	DehuiWu, Songling Huang,Wei Zhao & Junjun Xin. Infrared thermometer sensor dynamic error compensation using Hammerstein neural network. *Sensors and Actuators A*, 149 (2009), pp 152–158.
D0200	Qing-Yun Li and Yun-Hui Liu. A Combined Scheduling Scheme for Absolute and Relative Differentiated Services in Web-based Teleoperation. *Proceedings of the 2009 IEEE International Conference on Mechatronics and Automation*, August 2009.	O0227	J.-A. Fernández-Madrigal, C. Galindo, E. Cruz-Martín, A. Cruz-Martín & J. González. Automatic Regulation of the Information Flow in the Control Loops of a Web Teleoperated Robot. *Proceedings of the 2007 IEEE International Conference on Robotics and Automation* 2007, pp. 3496-3501
D0201	Mahesh Visvanathan, Adagarla, Bhargav Srinivas, Gerald Lushington, Sitta Sittapalam. Systematically Identifying Genes and Pathways in Multiple Cancer Types Using HGD & PSO-SVM. *Proceedings of the 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*, August 2009, pp. 494-497.	O0228	Nicolas Goffard & Georg Weiller. PathExpress: a web-based tool to identify relevant pathways in gene expression data. *Nucleic Acids Research*, 2007, Vol. 35, pp. 176-181.
		O0229	Li-Yeh Chuang, Cheng-Hong Yang, Chung-Jui Tu & and Cheng-Huei Yang. A Novel Feature Selection for Gene Expression Data. *Proceedings of the Joint Conference on Information Sciences*, Atlantis Press, October 2006
		O0230	Cheng-San Yang, Li-Yeh Chuang, Chao-Hsuan Ke & Cheng-Hong Yang. A Hybrid Feature Selection Method for Microarray Classification. *IAENG, International Journal of Computer Science*, 35-3, 2005

D0202	Jianfeng Zheng & Zaiqing Nie. Architecture and Implementation of an Object-level Vertical Search. *Proceedings of the International Conference on New Trends in Information and Service Science*, (NISS), June 2009, pp. 264-268.	O0231	Zaiqing Nie, Ji-Rong Wen & Wei-Ying Ma. Object-level Vertical Search. *3rd Biennial Conference on Innovative Data Systems Research (CIDR)* January 7-10 2007, Asilomar, California, USA.
D0203	Jianfeng Zheng & Zaiqing Nie. Language Models for Web Object Retrieval. *Proceedings of the International Conference on New Trends in Information and Service Science*, (NISS), June 2009, pp. 282-287.	O0232	Zaiqing Nie, Yunxiao Ma, Shuming Shi, Ji-Rong Wen & Wei-Ying Ma. *Web Object Retrieval*. Microsoft Research Technical Report: MSR-TR-2006-70.
D0204	Duggirala, R.; Gupta, R.; Zeng. Q-A; Agrawal, P. Performance Enhancements of Ad Hoc Networks with Localized Route Repair. *IEEE Transactions on Computers*, Vol. 52, No. 7, pp. 854-861, July 2003	O0233	Ionut¸ D. Aron and Sandeep K. S. Gupta. Analytical Comparison of Local and End-to-End Error Recovery in Reactive Routing Protocols for Mobile Ad Hoc Networks. *Proceedings of the 3rd ACM Int. Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM 2000)* pp. 69-76, August 2000.
D0205	M. Jalili-Kharaajoo, A. Rahmati and F. Rashidi. Internal Model Control Based on Locally Linear Model Tree (LOLIMOT) Model with Application to a PH Neutral Process. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics*, vol. 4, pp. 3051-3055	O0234	A. Fink, O. Nelles and R. Isermann, Nonlinear Internal Model Control for MISO Systems Based on Local Linear Neuro-Fuzzy Models. *15th IFAC World Congress*, Barcelona, Spain, 2002.
D0206	Hong Yin, Xian-Jia Wang, De-Bing Fang. A Novel Double Auction Mechanism for Electronic Commerce. *Proceedings of the 3rd International Conference on Machine Learning and Cybernetics*, August 2004	O0235	Peter Wurman, William Walsh, Michael Wellman. Flexible Double Auction for Electronic Commerce: Theory and Implementation. *Decision Support Systems*, Vol 24, Issue 1, November 1998, pp. 17-27.
D0208	Tianding Chen. A Novel Method for Protecting Sensitive Knowledge in Association Rules Mining. *2006 Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06)*	O0237	En Tzu Wang, Gualin Lee and Yu Tzu Lin. A Novel Method for Protecting Sensitive Knowledge in Association Rules Mining. *2005 Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSACC'05*)
D0209	Md. Maruf Monwar, Waqar Haque, Padma Polash Paul. A New Approach For Rotation Invariant Optical Character Recognition Using Eigendigit. *Canadian Conference on Electrical and Computer Engineering (CCECE 2007)*, 2007, pp. 1317 – 1320	O0238	Md. Al Mehedi Hasan, Md. Abdul Alim, Md. Wahedul Islam. A New Approach to Bangla Text Extraction and Recognition From Textual Image. *8th International Conference on Computer and Information Technology (ICCIT 2005)*, 2005, pp. 1 – 5.
D0210	Yong-Sung Choi; Ju-Ho Yun; Hyang-Kon Kim; Chung-Seog Choi; Kyung-Sup Lee. New Distribution Class Arrester Ground Lead Disconnector Design. *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, 21-24 April 2008, pp 731-73	O0240	Lenk D.W. New Distribution Class Arrester Ground Lead Disconnector Design Enhances Detonation Reliability and Improves Arrester Performance. *Proceedings of the 2004 Power Systems Conference and Exposition*, 10-13 Oct. 2004, vol. 3, pp. 1397-1406.

D0211	Huiling Wang. Business Blog Mining Based on Hierarchical SVM. *Proceedings of the 2008 International Symposium on Knowledge Acquisition and Modeling*, December 2008, pp 837-841.	O0241	Yun Chen, Flora S. Tsai and Kap Luk Chan. Machine Learning Techniques for Business Blog Search and Mining. *Expert Systems with Applications*, 35 (2008) October 2004, pp 581-590.
D0212	WenJie Tian, Yu Geng, Lan Ai, JiCheng Lu. Dynamic Compensation for Infrared Thermometer Based on SVR and Immune Clone Selection Algorithm. *International Conference on Industrial and Information Systems*, 2009, pp. 61–65	O0226	DehuiWu, Songling Huang,Wei Zhao & Junjun Xin. Infrared thermometer sensor dynamic error compensation using Hammerstein neural network. *Sensors and Actuators A*, 149 (2009), pp 152–158.
D0213	WenJie Tian, Yue Tian, Lan Ai, JiCheng Liu. Dynamic Compensation for Infrared Thermometer Based on Wiener Model and Particle Swarm Optimization Algorithm. *2nd IEEE International Conference on Computer Science and Information Technology (ICCSIT 2009)*, 2009, pp. 585–589	O0226	DehuiWu, Songling Huang,Wei Zhao & Junjun Xin. Infrared thermometer sensor dynamic error compensation using Hammerstein neural network. *Sensors and Actuators A*, 149 (2009), pp 152–158.
D0214	Chen Zhuolei. A Cooperative Method for Risk Management Based on Evolutionary Game Theory. *Proceedings of the ISECS International Colloquium on Computing, Communication, Control, and Management*, CCCM 2009, Vol.2, August 2009, pp. 344-347	O0243	D.J. Auda. Game Theory in Strategy Development of Reliability and Risk Management. *Proceedings of the Annual Reliability and Maintainability Symposium*, 2007. RAMS '07, January 2007, pp. 467-472
D0215	C. Unsihuay-Vila, J.W. Marangon-Lima, A.C. Zambroni de Souza. Integrated Operation and Expansion Planning of Natural Gas and Electricity Systems: Technical and Economical Aspects. *Proceedings of the 2009 IEEE PES General Meeting*, July 2009.	O0244	R. Rubio, D. Ojeda-Esteybar, O. Ano, A. Vargas. Integrated Natural Gas and Electricity Market: A Survey of the State of the Art in Operation Planning and Market Issues. *Proceedings of 2008 IEEE/PES Latin America Transmission and Distribution Conference and Expo*, August 2008.
D0216	Frank A. Ibikunle. Security Issues in Mobile WiMAX (IEEE 802.16e). *Proceedings of the Mobile WiMAX Symposium (MWS 2009)* July 2009, pp. 117-122.	O0245	Andreas Deininger,Shinsaku Kiyomoto, Jun Kurihara, Toshiaki Tanaka. Security Vulnerabilities and Solutions in Mobile WiMAX. *International Journal of Computer Science and Network Security*, Vol 7, No 11 November 2007.
D0217	Jun Yang and Zhigang Tang. The Exploration of Web Sites with Link Structure Graphs. *Proceedings of the 2009 International Conference on Management of eCommerce and eGovernment*, Sept. 2009, pp. 209-212.	O0246	Eduarda Mendes Rodrigues, Natasa Milic-Frayling, Martin Hicks, Gavin Smyth. Link Structure Graphs for Representing and Analyzing Web Sites. *Microsoft Research Technical Report MSR-TR-2006-94*, June 2006.
D0218	Ke Xu, Xiaoqi Zhang, Meina Song, Junde Song. Research on SLA Management Model in Service Operation Support System. *Proceedings of the 2009 International Conference on Wireless Communications Networking and Mobile Computing* (WiCom2009)	O0247	David Ameller and Xavier Franch. Service Level Agreement Monitor (SALMon). *Seventh International Conference on Composition-Based Software Systems, 2008*. (ICCBSS 2008), February 2008, pp.224-227

D0219	Jianfeng Zheng; Zaiqing Nie. Language Models for Web Object Retrieval. *Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, September 2009.	O0232	Zaiqing Nie, Yunxiao Ma, Shuming Shi, Ji-Rong Wen & Wei-Ying Ma. *Web Object Retrieval*. Microsoft Research Technical Report: MSR-TR-2006-70.
D0220	Asha Vijayan, Bessey Elen Skariah, Bipin G. Nair, Gerald H. Lushington, Sabarinath Subramanian, Mahesh Visvanathan. PathMapper-An Integrative Approach for Oncogene Pathway Identification. *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshop*, 2009. BIBMW 2009, November 2009, pp. 267-271	O0228	Nicolas Goffard & Georg Weiller. PathExpress: a web-based tool to identify relevant pathways in gene expression data. *Nucleic Acids Research*, 2007, Vol. 35, pp. 176-181.
D0221	T.M.N. Huda, S.F. Islam. Active Noise Control for Industrial Applications. *Proceedings of the First Asian Himalayas International Conference on Internet*, November 2009	O0191	Benny Sallberg, Lars Hakansson & Ingvar Claesson. Active Noise Control for Hearing Protection using a Low Power Fixed Point Digital Signal Processor. *Proceedings of the International Workshop on Acoustic Echo and Noise Control*, September 2005, pp. 65-68.
D0222	Liuzhou Wu; Zelin Chen. A Constraint-Based Text-to-Scene Conversion System. *Proceedings of the 2009 International Conference on Computational Intelligence and Software Engineering (CiSE 2009)*, December 2009	O0251	Bob Coyne, Richard Sproat. WordsEye: An Automatic Text-to-Scene Conversion System. *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, August 12, 2001, pp. 487-496
D0224	He Jing, Shi Dejia, and Wang Li. A Quadratic Particle Swarm Optimization for Weight Optimization. *Proceedings of the 2009 Third International Symposium on Intelligent Information Technology Application (IITA 2009)*, November 2009, pp. 557-560	O0253	Hui Li and Xuesong Yan. A New Optimization Algorithm for Weight Optimization. *Lecture Notes in Computer Science*, Volume 5370, Springer, 2008, pp. 723-730
D0225	Lei Wang, Baoyu Zheng, Jingwu Cui, Xiaorong Xu. Precoder Division Multiplexing Access for Cognitive Radio. *Proceedings of the International Conference on Wireless Communications & Signal Processing, 2009. WCSP 2009* , November 2009.	O0254	Leonard S. Cardoso, Mari Kobayashi, Oyvind Ryan, Merouane Debbah. Vandermonde Frequency Division Multiplexing for Cognitive Radio. *Proceedings of the IEEE 9th Workshop on Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008*, July 2008, pp. 421-425.
D0226	Tsao-Tsung Ma. Quantitative Design of Active Anti-Islanding Controllers for Power-Converter-Based Distributed Generators. *IEEE Transactions on Industrial Electronics*, Vol. 57, No. 10, 2010, pp. 3448–3455.	O0255	Seul-Ki Kim, Jin-Hong Jeon, Jong-Bo Ahn, Byongjun Lee, Sae-Hyuk Kwon. Frequency-Shift Acceleration Control for Anti-Islanding of a Distributed-Generation Inverter. *IEEE Transactions on Industrial Electronics*, Vol. 57, No. 2, February 2010, pp. 494–504.
		O0256	Seul-Ki Kim, Jin-Hong Jeon, Heung-Kwan Choi, Jonng-Bo Ahn. Design of Frequency Shift Acceleration Control for Anti-islanding of an Inverter-based DG. *13th Power Electronics and Motion Control Conference (EPE-PEMC 2008)*, 2008, pp. 2524–2529.

D0227	Dejia Shi, Li Wang, Jing He. Job Shop Scheduling Problem with an Novel Particle Swarm Optimization based on Tabu Search. *Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence*, Nov. 2009.	O0257	D.Y. Sha and Cheng-Yu Hsu. A Hybrid Particle Swarm Optimization for Job Shop Scheduling Problem. *Computers and Industrial Engineering*, Vol 51, No 4, Elsevier, December 2006, pp. 791-808.
D0228	Chen Yan and Wu Dan. A Scenario-Based Test Case Generation Framework for Security Policies. *International Conference on Computational Intelligence and Security, CIS '09*, December 2009, pp. 489-493.	O0258	Qaisar A. Malik, Johan Lilius and Linas Laibinis. Model-based Testing Using Scenarios and Event-B Refinements. *Methods, Models and Tools for Fault-Tolerance, Lecture Notes in Computer Science series*, Springer, March 2009, pp. 177-195
D0229	M. Jain and H. Kandwal. A Survey on Complex Wormhole Attack in Wireless Ad Hoc Networks. *Proceedings of the International Conference on Advances in Computing, Control, & Telecommunication Technologies*, 2009. ACT '09. December 2009, pp. 555-558.	O0259	Marianne A.Azer, Sherif M. El-Kassas, Magdy S. El-Soudani. A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks. *International Journal of Computer Science and Information Security*, Special Issue, May 2009.
D0230	Umesh Sehgal, Kuljeet Kaur, Pawan Kumar. Security in Vehicular Ad-hoc Networks. *Proceedings of the 2009 Second International Conference on Computer and Electrical Engineering ICCEE 09*, December 2009, pp. 485-488.	O0260	Arun Kumar Bayya, Siddhartha Gupte, Yogesh Kumar Shukla, Anil Garikapati. *Security in Ad-hoc Networks.* Student paper, Computer Science Department, University of Kentucky, 2006.
D0231	R. Sudha. A Tool for Identifying Ragas Using MIDI (Musical Instrument Devices) for CMIR (Classical Music Information Retrieval). *Proceedings of the 8th IEEE International Conference on Dependable, Autonomic and Secure Computing*, December 2009, pp. 837-842	O0261	Surendra Shetty and K.K. Achary. Raga Mining of Indian Music by Extracting Arohana-Avarohana Pattern. *International Journal of Recent Trends in Engineering*, Vol 1, No 1, May 2009, pp. 362-366.
D0232	S.K. Pal, S.S. Sarma, S.G. Neogi. A Pascal Graph Property and Computer Network Topology. *Proceedings of the First International Conference on Networks and Communications. NETCOM '09*, December 2009, pp.151-154	O0262	Somdev Chatterjee and Samar Sen Sarma. A New Property of Pascal Graph from the Perspective of Computer Network Topology. *International Journal of Computer Science and Information Technology (IJCSIT)*, Vol.1, 2007, pp. 14-20
D0233	WenJie Tian, JiCheng Liu. A New Optimization Algorithm for Dynamic Compensation of Sensors. *2010 Second International Conference on Computer Modeling and Simulation (ICCMS 2010)*, pp. 38–41.	O0226	DehuiWu, Songling Huang,Wei Zhao & Junjun Xin. Infrared thermometer sensor dynamic error compensation using Hammerstein neural network. *Sensors and Actuators A*, 149 (2009), pp 152–158.
D0234	T.T. Ma. Novel voltage stability constrained positive feedback anti-islanding algorithms for the inverter-based distributed generator systems. *IET Renewable Power Generation*, Vol. 4, No. 2, 2010, pp. 176–185.	O0255	Seul-Ki Kim, Jin-Hong Jeon, Jong-Bo Ahn, Byongjun Lee, Sae-Hyuk Kwon. Frequency-Shift Acceleration Control for Anti-Islanding of a Distributed-Generation Inverter. *IEEE Transactions on Industrial Electronics*, Vol. 57, No. 2, February 2010, pp. 494–504.

		O0256	Seul-Ki Kim, Jin-Hong Jeon, Heung-Kwan Choi, Jonng-Bo Ahn. Design of Frequency Shift Acceleration Control for Anti-islanding of an Inverter-based DG. *13th Power Electronics and Motion Control Conference (EPE-PEMC 2008)*, 2008, pp. 2524–2529.
D0235	Seyed Hassan Sadeghzadeh, Seyed Javad Mirabedini Shirazani, Mohammad Mosleh. A New Secure Scheme Purposed for Recognition and Authentication Protocol in Bluetooth Environment. *Proceedings of the 12th International Conference on Advanced Communication Technology (ICACT 2010)*, Feb. 2010	O0263	Mostafa Akhavan-E-Saffar, Vahid Tabataba Vakily. Improvement Bluetooth Authentication and Pairing Protocol Using Encrypted Key Exchange and Station to Station MAC Protocols. *Proceedings of the 2009 International Conference on Machine Learning and Computing (ICMLC 2009)*, July 2009, pp. 457-465.
D0236	Majid Shakeri and Ebrahim Hosseini. Simulation and Evaluation of Routing Protocols in Wireless Mobile Ad Hoc Networks. *2010 IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS)*, pp. 603 – 607.	O0264	Sung-Ju Lee, Julian Hsu, Russell Hayashida, Mario Gerla, Rajive Bagrodia. Selecting a routing strategy for your ad hoc network. *Computer Communications* Vol. 26, 2003, pp. 723 – 733.
D0237	Yong Zhang, Chen Wang. Design and Simulation of FIR Digital Filter using Fuzzy Adaptive Method. *Proceedings of the 2010 International Conference on Wireless Communications.*	O0265	Hime Aguiar e Oliveira, Antonio Petraglia & Mariane Rembold Petraglia. Frequency Domain FIR Filter Design Using Fuzzy Adaptive Simulated Annealing. *Circuits, Systems, and Signal Processing*, Vol 28, No 6, 2009, Elsevier pp. 899-911.
D0239	B. Noshad, M. Keramatzadeh, M. Saniei. Finding the Location of Switched Capacitor Banks in Distribution Systems Based on Wavelet Transform. *Proceedings of the 45th International Universities Power Engineering Conference (UPEC)*, August 2010.	O0266	H. Khani, M. Moallem & S. Sadri. On Tracking and Finding the Location of Switched Capacitor Banks in Distribution Systems. *Proceedings of the Transmission & Distribution Conference & Exposition:* Asia and Pacific, October 2009
		O0267	H. Khani, M. Moallem, S. Sadri. A Novel Algorithm for Determining the Exact Location of Switched Capacitor Banks in Distribution Systems. *Proceedings of the Transmission & Distribution Conference & Exposition: Asia and Pacific*, October 2009
D0240	Z. Yu and W. Wu. A Novel Optimization Method of Transceivers in Multiuser MIMO Wireless System. *Proceedings of the 2005 Second IEEE Consumer Communications and Networking Conference*, pp. 590-594.	O0269	S. Serbetli and A. Yener. Transceiver Optimization for Multiuser MIMO Systems. *IEEE Transactions on Signal Processing*, vol. 52, no. 1, January 2004, pp. 214-216.
D0241	Peng Zhao, Xuewu Cao, and Ping Luo. Attack on RADIUS Authentication Protocol. *2003 Proceedings of the International Conference on Communication Technology*, pp. 208-212	O0270	Joshua Hill. An Analysis of the RADIUS Authentication Protocol. *Proceedings of ICCT2003.*

D0242	Dong-Yu Xu, Guang-Rong Tang and Yu-Pin Luo. Dynamic Removal Algorithm for Constrained Delaunay Triangulations. *Proceedings of the Third International Conference on Image and Graphics*, 2004.	O0271	Marcelo Kallman, Hanspeter Bieri, and Daniel Thalmann. Fully Dynamic Constrained Delaunay Triangulations. *Geometric Modelling for Scientific Visualization*, G. Brunnett, B. Hamann, H.Mueller, L. Linsen (Eds.), Springer-Verlag, 2003, pp. 241-257.

Annexe 2.
Les 53 distances anormalement faibles dans le corpus O (α = 0.3%).

Rang	Texte A (Auteurs & references)	Texte B(Auteurs & references)	D(A,B)
1	O0054 : Thomas Weng, Yi Zhu & Chung-Kuan Cheng. Digital Design and Programmable Logic Boards: Do Students Actually Learn More? *38th ASEE/IEEE Frontiers in Education Conference*, October 22–25, 2008, Saratoga Springs, NY.	O0055 : Yi Zhu, Thomas Weng & Chung-Kuan Cheng. Enhancing Learning Effectiveness in Digital Design Courses Through the Use of Programmable Logic Boards. *IEEE Transactions on Education*, Vol. 52, N°. 1, February 2009.	0,1589
2	O0014 : M. Solanki, A. Cau & H. Zedan. Augmenting Semantic Web Service Description with Compositional Specifications. *Proceedings of the International World Wide Web Conference*, May 2004, pp. 544 - 552.	O0016 : M. Solanki, A. Cau & H. Zedan Introducing Compositionality in Web Service Descriptions. *Proceedings of the 10th IEEE International Workshop on Future Trends of Distributed Computing Systems* (FTDCS'04) 2004, pp 14-20..	0,2307
3	O0221 Hilario Gómez-Moreno, Saturnino Maldonado-Bascón & Francisco López-Ferreras. Edge Detection in Noisy Images Using the Support Vector Machines. *Lecture Notes in Computer Science*. Vol. 2084, pp. 685-692, 2001.	O0222 : H. Gómez Moreno, S. Maldonado Bascón, F. López Ferreras, P. Gil Jiménez. A New and Improved Edge Detector Using the Support Vector Machines. *Advances in Systems Engineering, Signal Processing and Communications*, 2002, pp. 239-243.	0,2446
4	O0019 : Santonu Sarkar, Avinash C. Kak & N. S. Nagaraja. Metrics for Analyzing Module Interactions in Large Software Systems. *Proceedings of the 12th Asia-Pacific Software Engineering Conference (APSEC?'05).*	O0020 : Santonu Sarkar, Girish Maskeri Rama & Avinash C. Kak. Modularization API-Based and Information-Theoretic Metrics for Measuring the Quality of Software *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, VOL. 33, NO. 1, JANUARY 2007.	0,2960
5	O0163 : M. Godoy Simoes & Bimal K. Bose. Fuzzy Logic Based Intelligent Control of a Variable Speed Cage Machine Wind Generation System. *IEEE Transactions on Power Electronics*, vol.12, no.1, Jan 1997, pp. 87-95.	O0164 : M. Godoy Simoes, Bimal K. Bose & Ronald J. SpiegelDesign and Performance Evaluation of a Fuzzy-logic-based Variable-speed Wind Generation System. *IEEE Transactions on Industry Applications*, vol.33, no.4, Jul/Aug 1997, pp. 956-965,.	0,2979
6	O0255 : Seul-Ki Kim, Jin-Hong Jeon, Jong-Bo Ahn, Byongjun Lee, Sae-Hyuk Kwon. Frequency-Shift Acceleration Control for Anti-Islanding of a Distributed-Generation Inverter. *IEEE Transactions on Industrial Electronics*, Vol. 57, No. 2, February 2010, pp. 494–504	O0256 : Seul-Ki Kim, Jin-Hong Jeon, Heung-Kwan Choi, Jonng-Bo Ahn. Design of Frequency Shift Acceleration Control for Anti-islanding of an Inverter-based DG. *13th Power Electronics and Motion Control Conference (EPE-PEMC 2008)*, pp. 2524–2529	0,3012
7	O0128 : Yuhua Li, Zuhair Bandar, David McLean & James O'Shea. A Method for Measuring Sentence Similarity and its Application to Conversational Agents. *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, pp. 820–825.	O0130 : Yuhua Li, David McLean, Zuhair A. Bandar, James D. O'Shea & Keeley Crockett. Sentence Similarity Based on Semantic Nets and Corpus Statistics. *IEEE Transactions on Knowledge and Data Engineering*, 2006, vol.18, no.8, pp.1138-1150.	0,3058
8	O0044 : Murat Yuksel, Shivkumar Kalyanaraman & Anuj Goel. Congestion Pricing Overlaid on Edge-to-Edge Congestion Control *IEEE International Conference on Communications* 2003, Vol. 2, 2003, 880 – 884.	O0045 : Murat Yuksel & Shivkumar Kalyanaraman. Distributed dynamic capacity contracting: an overlay congestion pricing framework. *Computer Communications* 26 (2003) 1484–1503.	0,3279
9	O0034 : S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part I: Concepts, Approaches, and Technical Challenges *IEEE Transactions on*	O0035 : S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part II: Technologies, Standards, and Tools for Building Multi-agent Systems.	0,3368

	Power Systems, Vol. 22 , No4, November 2007 , pp. 1743-1752.	*IEEE Transactions on Power Systems*, Vol. 22 , No.4 , November 2007 , pp. 1753-1759.	
10	O0080 : Md Mostofa Akbar, Eric G. Manning, Gholamali C. Shoja & Shahadat Khan. Heuristic Solutions for the Multiple-Choice Multi-Dimension Knapsack Problem. V.N. Alexandrov et al. (Eds.): *ICCS 2001*, LNCS 2074, pp. 659−668.	O0081 : Md Mostofa Akbar, M. Sohel Rahman, M. Kaykobad, E.G. Manning & G.C. Shoja. Solving the Multidimensional Multiple-choice Knapsack Problem by constructing convex hulls. *Computers & Operations Research* 33 (2006) 1259–1273	0,3467
11	O0048 : H. Khani, M. Moallem & S. Sadri. On Tracking and Finding the Location of Switched Capacitor Banks in Distribution Systems. *Proceedings of the Transmission & Distribution Conference & Exposition: Asia and Pacific*, October 2009	O0267 : H. Khani, M. Moallem, S. Sadri. A Novel Algorithm for Determining the Exact Location of Switched Capacitor Banks in Distribution Systems. *Proceedings of the Transmission & Distribution Conference & Exposition: Asia and Pacific*, October 2009	0,3482
12	O0018 : Victoria M. Catterson, Euan M. Davidson & Stephen D. J. McArthur. Issues in Integrating Existing Multi-agent Systems for Power Engineering Applications. *Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems*, Nov. 2005.	O0035 : S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part II: Technologies, Standards, and Tools for Building Multi-agent Systems. *IEEE Transactions on Power Systems*, Vol. 22.	0,3535
13	O0061 : Mathieu Jaume & Charles Morisset. Towards a formal specification of access control. *Proceedings of the Workshop on Foundations of Computer Security and Automated Reasoning for Security Protocol Analysis* (FCS-ARSPA'2006)	O0062 Mathieu Jaume & Charles Morisset A formal approach to implement access control models *Journal of Information Assurance and Security*, 2 (2006), pp. 59-70.	0,3585
14	O0229 : Li-Yeh Chuang, Cheng-Hong Yang, Chung-Jui Tu & and Cheng-Huei Yang. A Novel Feature Selection for Gene Expression Data. *Proceedings of the Joint Conference on Information Sciences*, Atlantis Press, October 2006	O0230 : Cheng-San Yang, Li-Yeh Chuang, Chao-Hsuan Ke & Cheng-Hong Yang. A Hybrid Feature Selection Method for Microarray Classification. *IAENG, International Journal of Computer Science*, 35-3, 2005	0,3599
15	O0231 Zaiqing Nie, Ji-Rong Wen & Wei-Ying Ma. Object-level Vertical Search. *3rd Biennial Conference on Innovative Data Systems Research* (CIDR) January 7-10 2007, Asilomar, California, USA.	O0232 : Zaiqing Nie, Yunxiao Ma, Shuming Shi, Ji-Rong Wen & Wei-Ying Ma Web. Object Retrieval. *Microsoft Research Technical Report*: MSR-TR-2006-70	0,3706
16	O0089 : John T. E. Timm & Gerald C. Gannod. A Model-Driven Approach for Specifying Semantic Web Services Proceedings of the 2005 IEEE International Conference on Web Services (ICWS 2005), vol.1, July 2005, pp. 313- 320.	O0273 Gerald C. Gannod & John T. E. Timm. An MDA-based Approach for Facilitating Adoption of Semantic Web Service Technology. *Proceedings of the IEEE EDOC Workshop on Model-Driven Semantic Web* (MDSW04), Sept. 2004, pp. 654-675.	0,3785
17	O0028: S. Khan, K. F. Li, E. G. Manning, M. M. Akbar. Solving the Knapsack Problem for Adaptive Multimedia Systems. *Studia Informatica*, 2002	O0080 : Md Mostofa Akbar, Eric G. Manning, Gholamali C. Shoja & Shahadat Khan. Heuristic Solutions for the Multiple-Choice Multi-Dimension Knapsack Problem. V.N. Alexandrov et al. (Eds.): *ICCS 2001*, LNCS 2074, pp. 659−668.	0,4018
18	O0024 : Chin-Yang Tseng, Poornima Balasubramanyam, Calvin Ko, Rattapon Limprasittiporn, Jeff Rowe & Karl Levitt. "A Specification-based Intrusion Detection System for AODVC". *Proceedings of the 2003 ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN-2003)*, Fairfax, VA USA, 31October 2003.	O0260 : Arun Kumar Bayya, Siddhartha Gupte, Yogesh Kumar Shukla, Anil Garikapati. *Security in Ad-hoc Networks*. Student paper, Computer Science Department, University of Kentucky, 2006.	0,4065
19	O0003 : Gregory M. Cooper, Eric A. Stone, George Asimenos, NISC Comparative	O0011 : Gregory M. Cooper, Michael Brudno, NISC Comparative Sequencing Program, Eric	0,4075

	Sequencing Program, Eric D. Green, Serafim Batzoglou and Arend Sidow. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, Jul 2005; 15, pp 901 – 913.	D. Green, Serafim Batzoglou, and Arend Sidow. Quantitative Estimates of Sequence Divergence for Comparative Analyses of Mammalian Genomes. *Genome Research*, May 2003; 13, pp 813 – 820.	
20	O0018 : Victoria M. Catterson, Euan M. Davidson & Stephen D. J. McArthur. Issues in Integrating Existing Multi-agent Systems for Power Engineering Applications. *Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems*, Nov. 2005.	O0034 : S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part I: Concepts, Approaches, and Technical Challenges *IEEE Transactions on Power Systems*, Vol. 22 , No4, November 2007 , pp. 1743-1752.	0,4077
21	O0086 : Yufei Wu, Brian D. Woerner & William J. Ebel. A Simple Stopping Criterion for Turbo Decoding. *IEEE COMMUNICATIONS LETTERS*, VOL. 4, NO. 8, AUGUST 2000	O0087 : Rose Y. Shao, Shu Lin & Marc P. C. Fossorier. Two Simple Stopping Criteria for Turbo Decoding. *IEEE TRANSACTIONS ON COMMUNICATIONS*, VOL. 47, NO. 8, AUGUST 1999, pp.1117-1120.	0,4185
22	O0259 : Marianne A.Azer, Sherif M. El-Kassas, Magdy S. El-Soudani. A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks. *International Journal of Computer Science and Information Sec*urity, Special Issue, May 2009	O0260 Arun Kumar Bayya, Siddhartha Gupte, Yogesh Kumar Shukla, Anil Garikapati. *Security in Ad-hoc Networks*. Student paper, Computer Science Department, University of Kentucky, 2006	0,4218
23	O0073 : John Penix & Perry Alexander. Toward Automated Component Adaptation. *Proceedings of the Ninth International Conference on Software Engineering and Knowledge Engineering* (SEKE-97), pp 535.542.	O0074 : John Penix. REBOUND: A Framework for Automated Component Adaptation. *Proceedings of the 9th Annual Workshop on Software Reuse* (WISR-99), 1999.	0,4232
24	O0130 : O0130 : Yuhua Li, David McLean, Zuhair A. Bandar, James D. O'Shea & Keeley Crockett. Sentence Similarity Based on Semantic Nets and Corpus Statistics. *IEEE Transactions on Knowledge and Data Engineering*, 2006, vol.18, no.8, pp.1138-1150.	O0131 : Aminul Islam & Diana Inkpen. Semantic Similarity of Short Texts. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, Bulgaria, September 2007.	0,4254
25	O0124 : Stephen D. J. McArthur, Scott M. Strachan & Gordon Jahn. The Design of a Multi-Agent Transformer Condition Monitoring System. *IEEE Transactions on Power Systems*, Vol 19, No 4, November 2004, pp. 1845-1852.	O0161 : V. M. Catterson & S. D. J. McArthur. The Practical Implications of Bringing a Multi-Agent Transformer Condition Monitoring System On-Line. *Proceedings of the 2004 Power Systems Conference and Exposition*, 10-13 Oct. 2004, vol. 1, pp.12 - 16.	0,4272
26	O0031 : Christophe Fumeaux, Dirk Baumann & Rüdiger Vahldieck. Finite-Volume Time-Domain Analysis of a Cavity-Backed Archimedean Spiral Antenna. *IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION*, VOL. 54, NO. 3, MARCH 2006	O0274 : C. Fumeaux, K. Sankaran, R. Vahldieck. Spherical Perfectly Matched Absorber for Finite-volume 3-D Domain Truncation. *IEEE Transactions on Microwave Theory and Techniques*, Vol. 55, No. 12, December 2007, pp. 2773-2781.	0,4325
27	O0028: S. Khan, K. F. Li, E. G. Manning, M. M. Akbar. Solving the Knapsack Problem for Adaptive Multimedia Systems. *Studia Informatica*, 2002	O0081 : Md Mostofa Akbar, M. Sohel Rahman, M. Kaykobad, E.G. Manning & G.C. Shoja. Solving the Multidimensional Multiple-choice Knapsack Problem by constructing convex hulls. *Computers & Operations Research* 33 (2006) 1259–1273	0,4337
28	O0018 : Victoria M. Catterson, Euan M. Davidson & Stephen D. J. McArthur. Issues in Integrating Existing Multi-agent Systems for Power Engineering Applications. *Proceedings*	O0124 : Stephen D. J. McArthur, Scott M. Strachan & Gordon Jahn. The Design of a Multi-Agent Transformer Condition Monitoring System. *IEEE Transactions on Power Systems*,	0,4351

	of the 13th International Conference on Intelligent Systems Application to Power Systems, Nov. 2005.	Vol 19, No 4, November 2004, pp. 1845-1852.	
29	O0024 : Chin-Yang Tseng, Poornima Balasubramanyam, Calvin Ko, Rattapon Limprasittiporn, Jeff Rowe & Karl Levitt. "A Specification-based Intrusion Detection System for AODVC". *Proceedings of the 2003 ACM Workshop on Security of Ad Hoc andSensor Networks (SASN-2003*), Fairfax, VA USA, 31October 2003.	O0259 Marianne A.Azer, Sherif M. El-Kassas, Magdy S. El-Soudani. A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks. *International Journal of Computer Science and Information Sec*urity, Special Issue, May 2009.	0,4356
30	O0018 : Victoria M. Catterson, Euan M. Davidson & Stephen D. J. McArthur. Issues in Integrating Existing Multi-agent Systems for Power Engineering Applications. *Proceedings of the 13th International Conference on Intelligent Systems Application to Power Systems*, Nov. 2005.	O0161 : V. M. Catterson & S. D. J. McArthur. The Practical Implications of Bringing a Multi-Agent Transformer Condition Monitoring System On-Line. *Proceedings of the 2004 Power Systems Conference and Exposition*, 10-13 Oct. 2004, vol. 1, pp.12 - 16.	0,4397
31	O0049 : Vik Tor Goh & Mohammad Umar Siddiqi. Multiple Error Detection and Correction Based on Redundant Residue Number Systems. *IEEE TRANSACTIONS ON COMMUNICATIONS*, VOL. 56, NO. 3, MARCH 2008	O0050 : Hari Krishna, Kuo-Yu Lin & Jenn-Dong Sun. A Coding Theory Approach to Error Control in Redundant Residue Number Systems. Part I: Theory and Single Error Correction. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol.39, no.1, Jan 1992, pp.8-17	0,4411
32	O0110 : Golnaz Karbaschi & Anne Fladenmuller. A Link-Quality and Congestion-aware Cross layer Metric for Multi-Hop Wireless Routing. *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference (MASS)*, Nov. 2005.	O0259 Marianne A.Azer, Sherif M. El-Kassas, Magdy S. El-Soudani. A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks. *International Journal of Computer Science and Information Sec*urity, Special Issue, May 2009	0,4491
33	O0233 : Ionut¸ D. Aron and Sandeep K. S. Gupta. Analytical Comparison of Local and End-to-End Error Recovery in Reactive Routing Protocols for Mobile Ad Hoc Networks. *Proceedings of the 3rd ACM Int. Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (MSWIM 2000) pp. 69-76, August 2000.	O0259 Marianne A.Azer, Sherif M. El-Kassas, Magdy S. El-Soudani. A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks. *International Journal of Computer Science and Information Sec*urity, Special Issue, May 2009.	0,4598
34	O0034 : S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part I: Concepts, Approaches, and Technical Challenges. *IEEE Transactions on Power Systems*, Vol. 22 , No4, November 2007 , pp. 1743-1752.	O0124 : Stephen D. J. McArthur, Scott M. Strachan & Gordon Jahn. The Design of a Multi-Agent Transformer Condition Monitoring System. IEEE Transactions on Power Systems, Vol 19, No 4, November 2004, pp. 1845-1852.	0,4602
35	O0120 : Feifei Gao & A. Nallanathan. Blind Channel Estimation for MIMO OFDM Systems via Nonredundant Linear Precoding. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, VOL. 55, NO. 2, FEBRUARY 2007	O0269 : S. Serbetli and A. Yener Transceiver Optimization for Multiuser MIMO Systems. *IEEE Transactions on Signal Processing*, vol. 52, no. 1, January 2004, pp. 214-216.	0,4615
36	O0110 : Golnaz Karbaschi & Anne Fladenmuller. A Link-Quality and Congestion-aware Cross layer Metric for Multi-Hop Wireless Routing. *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference (MASS)*, Nov. 2005	O0233 : Ionut¸ D. Aron and Sandeep K. S. Gupta. Analytical Comparison of Local and End-to-End Error Recovery in Reactive Routing Protocols for Mobile Ad Hoc Networks. *Proceedings of the 3rd ACM Int. Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (MSWIM 2000) pp. 69-76,	0,4623

		August 2000.	
37	O0013 : Mattias Nordin & Per-Olof Gutmanin. Controlling mechanical systems with backlash - a survey. *Automatica.*38 (2002), pp. 1633–1649.	O0021 : P. Rostalski, T. Besselmann, M. Bari, F. Van Belzen & M. Morari. A hybrid approach to modelling, control and state estimation of mechanical systems with backlash. *International Journal of Control*, vol. 80, no. 11, pp. 1729-1740.	0,4629
38	O0035 : S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part II: Technologies, Standards, and Tools for Building Multi-agent Systems. *IEEE Transactions on Power Systems*, Vol. 22	O0124 : Stephen D. J. McArthur, Scott M. Strachan & Gordon Jahn. The Design of a Multi-Agent Transformer Condition Monitoring System. *IEEE Transactions on Power Systems*, Vol 19, No 4, November 2004, pp. 1845-1852.	0,4629
39	O0010 : Sergey Brin, Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30 (1998), p 107- 117.	O0078 : Filippo Menczer, Ruj Akavipat & Le-Shin Wu6S: Distributing crawling and searching across Web peers. *CIKM 2004* Washington, D.C. USA.	0,4688
40	O0102 : Sugata Sanyal, Ajith Abraham, Dhaval Gada, Rajat Gogri, Punit Rathod, Zalak Dedhia & Nirali Mody. Security Scheme for Distributed DoS in Mobile Ad Hoc Networks. *Sixth International Workshop on Distributed Computing* (IWDC), vol. 3326, LNCS. Springer, 2004.	O0259 Marianne A.Azer, Sherif M. El-Kassas, Magdy S. El-Soudani. A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks. *International Journal of Computer Science and Information Sec*urity, Special Issue, May 2009	0,4690
41	O0130 : Yuhua Li, David McLean, Zuhair A. Bandar, James D. O'Shea & Keeley Crockett. Sentence Similarity Based on Semantic Nets and Corpus Statistics.*IEEE Transactions on Knowledge and Data Engineering*, 2006, vol.18, no.8, pp.1138-1150	O0194 : Lucy Vanderwendea, Hisami Suzukia, Chris Brocketta & Ani Nenkova. Beyond SumBasic: Task-Focused Summarization with Sentence Simplification and Lexical Expansion. *Information Processing and Management*, Vol 43, Elsevier, 2007, pp 1606-1618.	0,4704
42	O0010 : Sergey Brin, Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30 (1998), p 107- 117.	O0231 : Zaiqing Nie, Ji-Rong Wen & Wei-Ying Ma. Object-level Vertical Search. *3rd Biennial Conference on Innovative Data Systems Research* (CIDR) January 7-10 2007, Asilomar, California, USA.	0,4718
43	O0219 : H. Boudali & J.B. Dugan A discrete-time Bayesian network reliability modeling and analysis framework. *Reliability Engineering and System Safety*, 87, March 2005, pp. 337–349.	O0225 : Tariq Assaf & Joanne Bechta Dugan. Diagnostic Expert Systems from Dynamic Fault Trees. *Proceedings of the 2004 Annual Symposium Reliability and Maintainability*, Jan. 2004, pp. 444- 450.	0,4722
44	O0110 : Golnaz Karbaschi & Anne Fladenmuller. A Link-Quality and Congestion-aware Cross layer Metric for Multi-Hop Wireless Routing. *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference (MASS)*, Nov. 2005	O0264 : Sung-Ju Lee, Julian Hsu, Russell Hayashida, Mario Gerla, Rajive Bagrodi Selecting a routing strategy for your ad hoc network.*Computer Communications* Vol. 26, 2003, pp. 723 – 733.	0,4751
45	O0202 : John Stang. *The Inverted Pendulum Design Project Report*, Cornell University, May 2005.	O0203 : Jordan Crittenden & Parker Evans. *MEMS Inertial Navigation System*. Student Project: ECE4910, Cornell University, May 8, 2008.	0,4753
46	O0078 : Filippo Menczer, Ruj Akavipat & Le-Shin Wu. 6S: Distributing crawling and searching across Web peers. *CIKM 2004* Washington, D.C. USA	O0231 : Zaiqing Nie, Ji-Rong Wen & Wei-Ying Ma. Object-level Vertical Search. *3rd Biennial Conference on Innovative Data Systems Research* (CIDR) January 7-10 2007, Asilomar, California, USA.	0,4761
47	O0134 : Gustavo M. D. Vieira & Luiz E. Buzato. Distributed Checkpointing: Analysis and Benchmarks. *Proceedings of the 24th Brazilian Symposium on Computer Networks*,	O0154 : Awadhesh Kumar Singh. On Mobile Checkpointing using Index and Time Together. *Proceedings of World Academy of Science, Engineering and Technology*, Vol 26, December	0,4771

	May 2006.	2007, pp.144-151.	
48	O0017 : Mario Vasak, Mato Baoti´c, Ivan Petrovi´c & Nedjeljko Peri´c. Hybrid Theory Based Time-Optimal Control of an Electronic Throttle.*IEEE-ISIE 2005*, June 2005, pp. 20-23.	O0021 : P. Rostalski, T. Besselmann, M. Bari, F. Van Belzen & M. Morari. A hybrid approach to modelling, control and state estimation of mechanical systems with backlash. *International Journal of Control*, vol. 80, no. 11, pp. 1729-1740	0,4775
49	O0128 : Yuhua Li, Zuhair Bandar, David McLean & James O'Shea A Method for Measuring Sentence Similarity and its Application to Conversational Agents. *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference* (FLAIRS 2004), pp. 820–825.	O0131 : Aminul Islam & Diana Inkpen. Semantic Similarity of Short Texts. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*, Bulgaria, September 2007	0,4796
50	O0259 : Marianne A.Azer, Sherif M. El-Kassas, Magdy S. El-Soudani. A Full Image of the Wormhole Attacks Towards Introducing Complex Wormhole Attacks in Wireless Ad Hoc Networks. *International Journal of Computer Science and Information Security*, Special Issue, May 2009	O0264 : Sung-Ju Lee, Julian Hsu, Russell Hayashida, Mario Gerla, Rajive Bagrodi. Selecting a routing strategy for your ad hoc network.*Computer Communications* Vol. 26, 2003, pp. 723 – 733.	0,4797
51	O0023 : M. Pipattanasomporn, H. Feroze & S. Rahman . Multi-agent Systems in a Distributed Smart Grid: Design and Implementation. *Proceedings of the IEEE/PES Power Systems Conference and Exposition*, 2009, pp.1-8.	O0035 : S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci & T. Funabashi. Multi-Agent Systems for Power Engineering Applications-Part II: Technologies, Standards, and Tools for Building Multi-agent Systems. *IEEE Transactions on Power Systems*, Vol. 22 , No.4 , November 2007 , pp. 1753-1759.	0,4800
52	O0051 : Christos A. Kontovas & Harilaos N. Psaraftis. Formal Safety Assessment: A Critical Review. *Marine Technology*, Vol. 46, No. 1, January 2009, pp. 45–59	O0243 : D.J. Auda. Game Theory in Strategy Development of Reliability and Risk Management. *Proceedings of the Annual Reliability and Maintainability Symposium*, 2007. RAMS '07, January 2007, pp. 467-472	0,4802
53	O0202 : John Stang. *The Inverted Pendulum* Design Project Report, Cornell University, May 2005.	O0218 : Chee Pin Tant & Maki K. Hahibt. Robust sensor fault reconstruction for an inverted pendulum using right eigenstructure assignment. *Proceedings of the 2004 IEEE Intemational Conference on Control Applications*. Taipei. Taiwan, September 2004, pp. 1236- 1241.	0,4823

Annexe 3

Classification des duplications (D) en fonction de leur(s) plus proche(s) voisin(s) originaux (O) avec leurs distances respectives (D(d,o))

D Duplicat	O1 Premier voisin	D(D,O1)	O2 Deuxième voisin	D(D,O2)	O3 Troisième voisin	D(D,O3)	Commentaire
D0001	O0005	0,422	O0004	0,431	O0007	0,434	Ok
D0002	O0011	0,376	O0003	0,461			OK
D0003	O0012	0,411					OK
D0004	O0002	0,234					OK
D0005	O0001	0,392					OK
D0006	O0009	0,349					OK
D0007	O0010	0,312					OK
D0008	O0024	0,323	O0260	0,437	O0259	0,450	OK
D0009	O0016	0,344	O0014	0,387			OK
D0010	O0020	0,302	O0019	0,308			OK
D0011	O0021	0,403	O0017	0,430	O0013	0,485	OK
D0012	O0015	0,163					OK
D0013	O0022	0,133					OK
D0014	O0018	0,168	O0035	0,376	O0034	0,434	OK
D0015	O0132	0,537					Hors limite
D0016	O0026	0,133					OK
D0017	O0027	0,034					OK
D0018	O0049	0,527					Hors limite
D0019	O0029	0,262					OK
D0020	O0030	0,197					OK
D0021	O0274	0,328	O0031	0,470			OK
D0022	O0032	0,358					OK
D0023	O0033	0,268					OK
D0024	O0035	0,447	O0034	0,455	O0018	0,486	OK
D0025	O0036	0,364					OK
D0026	O0037	0,134					OK
D0027	O0038	0,316					OK
D0028	O0039	0,085					OK
D0029	O0040	0,257					OK
D0030	O0041	0,204					OK
D0032	O0266	0,314					OK
D0033	O0044	0,346	O0045	0,351			OK
D0034	O0046	0,153					OK
D0035	O0047	0,489					OK
D0037	O0049	0,519					OK
D0038	O0051	0,325					OK
D0039	O0052	0,250					OK

D0040	O0053	0,289					OK
D0041	O0055	0,463	O0054	0,482			OK
D0042	O0052	0,296					OK
D0043	O0056	0,149					OK
D0045	O0058	0,380					OK
D0046	O0059	0,364	O0060	0,400			OK
D0047	O0061	0,372	O0062	0,397			OK
D0048	O0063	0,195					OK
D0049	O0064	0,073					OK
D0050	O0065	0,446					OK
D0051	O0067	0,483					OK
D0052	O0068	0,435					OK
D0053	O0069	0,196					OK
D0054	O0070	0,385					OK
D0055	O0071	0,259					OK
D0056	O0072	0,353					OK
D0057	O0074	0,378	O0073	0,380			OK
D0058	O0075	0,172					OK
D0059	O0165	0,539					Hors limite
D0060	O0074	0,409	O0073	0,434	O0220	0,486	OK
D0061	O0077	0,326					OK
D0062	O0078	0,378	O0110	0,488	O0259	0,500	OK
D0063	O0079	0,304					OK
D0064	O0028	0,371	O0080	0,393	O0081	0,416	OK
D0065	O0083	0,495					Hors limite
D0066	O0084	0,215					OK
D0067	O0078	0,406					OK
D0068	O0085	0,259					OK
D0069	O0078	0,385	O0110	0,479	O0134	0,497	OK
D0070	O0086	0,470	O0087	0,477			OK
D0071	O0078	0,358	O0110	0,483	O0134	0,489	OK
D0072	O0088	0,408					OK
D0073	O0028	0,387	O0080	0,406	O0081	0,422	OK
D0074	O0089	0,126	O0273	0,367			OK
D0075	O0090	0,262					OK
D0076	O0091	0,322					OK
D0077	O0092	0,407					OK
D0078	O0105	0,289					OK
D0079	O0094	0,336					OK
D0080	O0095	0,197					OK
D0081	O0096	0,448					OK
D0082	O0097	0,527					Hors limite
D0083	O0098	0,363					OK

D0084	O0099	0,380					OK
D0085	O0100	0,121					OK
D0086	O0101	0,433					OK
D0087	O0102	0,364	O0259	0,426	O0024	0,459	OK
D0089	O0078	0,404					OK
D0090	O0104	0,159					OK
D0091	O0093	0,241					OK
D0092	O0106	0,384					OK
D0093	O0107	0,205					OK
D0094	O0108	0,411					OK
D0095	O0109	0,234					OK
D0096	O0110	0,275	O0259	0,469	O0233	0,477	OK
D0097	O0103	0,240					OK
D0098	O0111	0,426					OK
D0099	O0112	0,326					OK
D0100	O0113	0,282					OK
D0101	O0114	0,213					OK
D0102	O0115	0,255					OK
D0103	O0116	0,396	O0117	0,448			OK
D0105	O0119	0,305					OK
D0106	O0120	0,371					OK
D0107	O0121	0,307					OK
D0108	O0122	0,313	O0238	0,490	O0192	0,490	OK
D0109	O0123	0,336					OK
D0110	O0124	0,407	O0035	0,467	O0034	0,468	OK
D0111	O0125	0,348					OK
D0112	O0126	0,276					OK
D0113	O0127	0,217					OK
D0114	O0130	0,309	O0128	0,384	O0131	0,452	OK
D0115	O0132	0,493					Limite
D0116	O0078	0,437					OK
D0117	O0134	0,431	O0154	0,490			OK
D0118	O0112	0,216					OK
D0119	O0113	0,321					OK
D0120	O0136	0,278					OK
D0121	O0138	0,375	O0137	0,404			OK
D0122	O0140	0,350	O0139	0,383	O0141	0,384	OK
D0123	O0142	0,363					OK
D0124	O0143	0,206					OK
D0125	O0144	0,272					OK
D0126	O0145	0,061					OK
D0127	O0148	0,359	D0001	0,459	O0146	0,484	OK
D0128	O0150	0,227					OK

D0129	O0151	0,082					OK
D0130	O0152	0,317					OK
D0131	O0153	0,433					OK
D0132	O0154	0,032	O0134	0,479			OK
D0133	O0155	0,221					OK
D0134	O0156	0,150					OK
D0135	O0157	0,201					OK
D0136	O0158	0,236					OK
D0137	O0159	0,218					OK
D0138	O0160	0,159					OK
D0139	O0161	0,104	O0124	0,446	O0018	0,447	OK
D0140	O0162	0,465					OK
D0141	O0163	0,314	O0164	0,378			OK
D0142	O0165	0,472					OK
D0143	O0166	0,363					OK
D0144	O0167	0,121					OK
D0145	O0168	0,190					OK
D0146	O0169	0,416					OK
D0147	O0170	0,188					OK
D0148	O0171	0,407					OK
D0149	O0172	0,334					OK
D0150	O0173	0,388					OK
D0151	O0174	0,308					OK
D0152	O0175	0,378					OK
D0153	O0176	0,287					OK
D0154	O0177	0,508					Hors limite
D0155	O0178	0,082					OK
D0156	O0179	0,165					OK
D0157	O0180	0,267					OK
D0158	O0181	0,065					OK
D0159	O0182	0,360					OK
D0160	O0142	0,420					OK
D0161	O0184	0,208					OK
D0162	O0056	0,153					OK
D0163	O0186	0,242					OK
D0164	O0187	0,465					OK
D0165	O0188	0,032					OK
D0166	O0189	0,143					OK
D0167	O0190	0,124					OK
D0168	O0191	0,340					OK
D0169	O0192	0,362					OK
D0170	O0193	0,339	O0022	0,461	O0109	0,491	OK
D0171	O0194	0,480					OK

D0172	O0196	0,177					OK
D0173	O0197	0,018					OK
D0174	O0198	0,407	O0199	0,452	O0200	0,476	OK
D0175	O0190	0,313					OK
D0176	O0201	0,078					OK
D0177	O0202	0,323					OK
D0178	O0203	0,347					OK
D0179	O0204	0,233					OK
D0180	O0205	0,263					OK
D0181	O0206	0,108					OK
D0182	O0207	0,051					OK
D0183	O0208	0,204					OK
D0184	O0209	0,132					OK
D0185	O0210	0,377	O0211	0,475			OK
D0186	O0212	0,025					OK
D0187	O0213	0,281					OK
D0188	O0214	0,316					OK
D0189	O0215	0,403					OK
D0190	O0216	0,171					OK
D0191	O0217	0,314					OK
D0192	O0218	0,344					OK
D0193	O0219	0,460	O0224	0,490			OK
D0194	O0220	0,415					OK
D0195	O0222	0,240	O0221	0,241			OK
D0196	O0223	0,153					OK
D0197	O0225	0,455	O0224	0,479			OK
D0198	O0213	0,284					OK
D0199	O0226	0,341					OK
D0200	O0072	0,510					Hors limite
D0201	O0229	0,372	O0230	0,432			OK
D0202	O0231	0,300	O0232	0,453			OK
D0203	O0232	0,240	O0231	0,450			OK
D0204	O0233	0,375	O0110	0,445	O0260	0,446	OK
D0205	O0234	0,337					OK
D0206	O0235	0,347					OK
D0208	O0237	0,211					OK
D0209	O0238	0,495					Limite
D0210	O0240	0,127					OK
D0211	O0241	0,471					OK
D0212	O0226	0,360					OK
D0213	O0226	0,355					OK
D0214	O0243	0,218					OK
D0215	O0244	0,209					OK

D0216	O0245	0,131	O0091	0,484	O0260	0,487	OK
D0217	O0246	0,230					OK
D0218	O0247	0,432					OK
D0219	O0232	0,239	O0231	0,450			OK
D0220	O0228	0,341					OK
D0221	O0191	0,339					OK
D0222	O0251	0,381					OK
D0224	O0253	0,276					OK
D0225	O0254	0,335					OK
D0226	O0255	0,344	O0256	0,393			OK
D0227	O0257	0,350					OK
D0228	O0258	0,427					OK
D0229	O0259	0,267	O0260	0,474	O0024	0,489	OK
D0230	O0260	0,386					OK
D0231	O0261	0,375					OK
D0232	O0262	0,367					OK
D0233	O0226	0,350					OK
D0234	O0255	0,365	O0256	0,396			OK
D0235	O0263	0,194					OK
D0236	O0264	0,389					OK
D0237	O0265	0,299					OK
D0239	O0267	0,357	O0048	0,363			OK
D0240	O0269	0,499					OK
D0241	O0270	0,131					OK
D0242	O0271	0,393					OK

Annexe 4

Un exemple de détection des passages dupliqués avec la fenêtre glissante

Duplication	**Original**
(D0002 : du 1000^{e} au 1500^{e} mot) [armadillo, horse, cow, sheep, indianmunjtak, pig, rabit, galago, lemur, mouse-lemur, marmoset, dusky-titi, squirrel-monkey, vervet, baboon, macaque, oraguntam, gorilla, chimp, wallaby, monodelphis, opussum. The reference human sequence for this targeted region corresponds to NCBI build 35, i.e., human chromosome 7, 115404472—117281897. For all of our analyses] we treat the first human base in this region as position 1. This region contains 10 RefSeq genes, 40.2% repetitive DNA, and 38.4% G+C. Gene annotations for the human sequence were obtained from the UCSC Genome Browser (http://genome.ucsc.edu/) using the RefSeq gene track ([7]); this includes 151 unique exons (in which an exon consisting of both UTR and coding sequence is split into separate "unique" exons) totaling 36,959 bases. **Alignment** We used a combination of both global and local techniques to construct a multiple sequence alignment of set of these sequences. This strategy ensures that rearrangement events, identified as high-scoring local alignments, are properly captured and placed in the context of a global alignment. First, we compared each nonhuman sequence to the human using the program Shuffle-LAGAN ([8]). Shuffle-LAGAN is effective at the identification of rearrangements, such as translocations and inversions, in the context of a global, pairwise alignment. The nonhuman sequences are subsequently reordered and reoriented (i.e., shuffled) so that the local alignment chains are monotonic with respect to the human sequence. In this process, regions of the nonhuman sequences that lack detectable similarity to any region of the human are clipped and deleted. These rearranged sequences are thus orthologously collinear with the human sequence. We then aligned the rearranged sequences using MLAGAN, a global multiple sequence aligner previously shown to be effective and accurate for multiple alignment of mammalian genomic sequences ([9]). The tree supplied to MLAGAN for this step is similar to the topology shown in Figure 1(on the following page), but it includes a small number of topology changes designed to align longer branch groups later; this step is necessary because alignment accuracy is best when species with the greatest sequence similarity are aligned first. We used parameters similar to those used to generate alignments of the human, mouse, and rat genome	(O0003 : du 7250^{e} au 7750^{e} mot) we treat the first human base in this region as position 1. This region contains 10 RefSeq genes, 40.2% repetitive DNA, and 38.4% G+C. Gene annotations for the human sequence were obtained from the UCSC Genome Browser (http://genome.ucsc.edu/) using the RefSeq gene track ([7]); this includes 151 unique exons (in which an exon consisting of both UTR and coding sequence is split into separate "unique" exons) totaling 36,959 bases. **Alignment** We used a combination of both global and local techniques to construct a multiple sequence alignment of set of these sequences. This strategy ensures that rearrangement events, identified as high-scoring local alignments, are properly captured and placed in the context of a global alignment. First, we compared each nonhuman sequence to the human using the program Shuffle-LAGAN ([8]). Shuffle-LAGAN is effective at the identification of rearrangements, such as translocations and inversions, in the context of a global, pairwise alignment. The nonhuman sequences are subsequently reordered and reoriented (i.e., shuffled) so that the local alignment chains are monotonic with respect to the human sequence. In this process, regions of the nonhuman sequences that lack detectable similarity to any region of the human are clipped and deleted. These rearranged sequences are thus orthologously collinear with the human sequence. In this process, regions of the nonhuman sequences that lack detectable similarity to any region of the human are clipped and deleted. These rearranged sequences are thus orthologously collinear with the human sequence. We then aligned the rearranged sequences using MLAGAN, a global multiple sequence aligner previously shown to be effective and accurate for multiple alignment of mammalian genomic sequences (Brudno et al. 2003a). The tree supplied to MLAGAN for this step is similar to the topology shown in Figure 1B, but is rooted on the marsupial branch and includes a small number of topology changes designed to align longer branch groups later; this step is necessary because alignment accuracy is best when species with the greatest sequence similarity are aligned first. We used parameters similar to those used to generate alignments of the human, mouse, and rat genome sequences (Brudno et al. 2004).

sequences ([10]). Entire tree construction and estimation of the neutral rate for the entire tree. We extracted all of those regions from the uncompressed alignment corresponding to the highest-scoring constrained elements in the human sequence yielding an alignment of 97,274 columns. Using a species topology as shown in Figure 1 defined ([11];[12]), we obtained the maximum likelihood branch lengths using PHYLIP (http://evolution.genetics.washington.edu/phylip.html), with the HKY 85 model of nucleotide substitution ([13]). The ML-tree is shown in Figure 1. Given the relative branch-length tree (Figure 1), we estimated the neutral rate for the entire tree as follows. Briefly, we estimate the neutral divergence among closely related species (ranging from 3% to 10% difference; Table 1), and subsequently extrapolated these rate estimates over the entire relative branch-length tree. As a source of aligned neutral DNA, we began with the uncompressed global alignment and excluded all of those alignment regions containing unambiguously constrained elements in the human sequence (the complement of the alignment used to determine the relative branch-length tree	**Tree construction and estimation of the neutral rate** We extracted all of those regions from the uncompressed alignment corresponding to the highest-scoring constrained elements in the human sequence yielding an alignment of 97,274 columns. Using a species topology previously defined (Madsen et al. 2001; Murphy et al. 2001), we obtained the maximum likelihood branch lengths using SEMPHY (Friedman et al. 2002), with the HKY 85 model of nucleotide substitution (Hasegawa et al. 1985; Fig. 1B). Given the relative branch-length tree (Fig. 1B), we estimated the neutral rate for the entire tree essentially as previously described (Cooper et al. 2003). Briefly, we estimate the neutral divergence among closely related species (ranging from 3% to 10% difference; Table 3), and subsequently extrapolated these rate estimates over the entire relative branch-length tree. As a source of aligned neutral DNA, we began with the uncompressed global alignment and excluded all of those alignment regions containing unambiguously constrained elements in the human sequence (the complement of the alignment used to determine the relative branch-length tree above). Divergence estimates were then made for each closely related group of species (neutral rate between 0.03 and 0.10 subs/site) using baseml of the PAML (Yang 1997) software package with the HKY 85 model of nucleotide substitution (Hasegawa et al. 1985).

D0002 : Anup Bhatkar & J.L. Rana. Estimating neutral divergence amongst Mammals for Comparative Genomics with Mammalian Scope. *Proceedings of the 9th International Conference on Information Technology (ICIT'06).*

O0003 : Gregory M. Cooper, Eric A. Stone, George Asimenos. NISC Comparative Sequencing Program, Eric D. Green, Serafim Batzoglou and Arend Sidow, Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, Jul 2005; 15, pp 901 – 913.