IDEA: SBIR Phase I: #1620992

EA4 Educational Technologies and Applications SBIR Phase I:MyPatentIdeas.com -- From Search to Research with Fast Patent-document Correlations SBIR Phase I Proposal: From search to research with fast document correlations Problem Statement (Intellectual Merit) This Small Business [...]

(7591 total words; 1609 unique words; 1320 useful unique words)

[...] computing resources. We have the knowhow. We have the ambition. We now need the funding to make this happen. With the appropriate funds, we will be able to accelerate the research and commercialization process to make MyPatentIdeas.com a successful and valid patent-search service.

Key Phrases In IDEA: small business innovation research phase (1), patent documents represent prior art (1), small business innovation researcher (2), natural language processing nlp (1), pca principal components analysis (1), latent semantic analysis lsa (1), latent semantic index lsi (1), natural language processing methods (1), lsa latent semantic analysis (1), small business owner (4), term document matrix (4), patent document correlations (4), nearest neighbor patent (4), small business innovation (3), patent search service (3), smart search technologies (3), entire patent collection (3), patent documents (21), document correlations (15), patent data (12), natural language (11), query document (11), patent search (10), inputed idea (10), search service (9), patent rank (9), patent (170), search (119), document (108), terms (60), idea (57), concepts (41), technology (36)

SUBSET

In 252 seconds, 35 independent searches were performed on the COLLECTION, consisting of 2,128 phrases, 6,584 terms, and 80,782 internal queries producing a result set of **1,343** documents; that is, the SUBSET.

Key Phrases In SUBSET: field programmable gate array fpga (231), programmable data processing apparatus create (157), data processing apparatus create means (157), special purpose hardware based systems (144), application specific integrated circuit asic (173), public switched telephone network pstn (94), random access memory ram (577), local area network lan (713), wide area network wan (545), natural language processing nlp (1639), field programmable gate arrays (307), personal digital assistant pda (268)

The SUBSET consists of a total of 1,768,310 phrases. After deep scrutiny of this SUBSET using drill-down, statistical analyses, we identify the **100** most relevant documents; that is, the NEIGHBORHOOD.
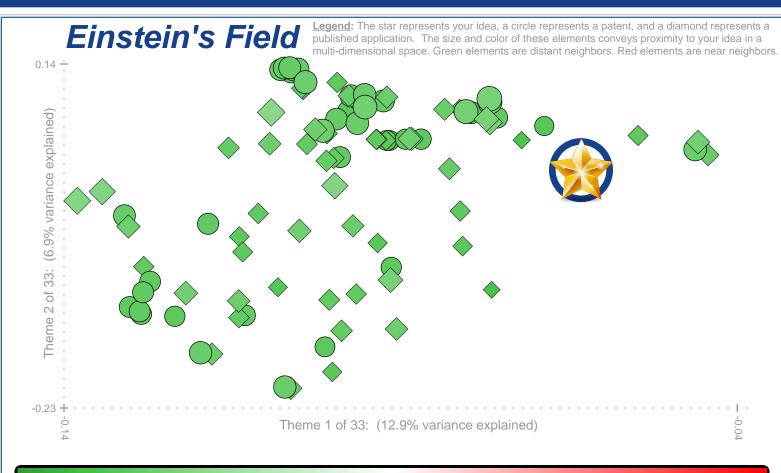
NEIGHBORHOOD

Top Technologies	[0.124 - diversified]
G06F	Physics => Computing => Electrical Digital Data Processing
G06Q	Physics => Computing => Data Processing Systems Or Methods, Specially Adapted For Administrative, Commercial, Financial, Managerial, Supervisory Or Forecasting Purposes
H04N	Electricity => Electric Communication Technique => Pictorial Communication, E.g. Television

Top Firms [0.042 - diversified]	Top Inventors [0.011 - diversified]
Digimarc Corporation	Bishop; Michael
Rose Blush Software LLC	Short; Shannon M.
AT&T Intellectual Property I, L.P.	McLendon; Martin L.
Accenture LLP	Beckham; Carol T.
THOMSON REUTERS GLOBAL RESOURCES	Frank; Scott M.
PATTERSON THUENTE PEDERSEN, P.A.	Rhoads; Geoffrey B.
VCVC III LLC	Rivette; Kevin G.
General Electric Company	Hohmann; Luke
Aurigin Systems, Inc.	Rappaport; Irving S.
FUJI XEROX CO., LTD.	Tran; Bao
INTERNATIONAL BUSINESS MACHINES CORPORATION	Lundberg; Steven W.
Universidad Politecnica de Catalunya	Jackson; Adam
Nortel Networks Limited	Thornthwaite; Warren

Einstein's Field

Theme 2 of 33: (6.9% variance explained)

Theme 1 of 33: (12.9% variance explained)

0.14

-0.23

-0.14

-0.04

DISTANT NEIGHBOR   NEAREST NEIGHBOR

P1 E1 N1

Rank	Sphinx	Pythagoras	Newton	Einstein
1	APP 20130173249	628▲ APP 20070136116 (P1)	APP 20070136116 (N1)	4▲ APP 20150310114 (E1): "METHOD, SYSTEM AND SOFTWARE FOR SEARCHING, IDENTIFYING, RETRIEVING AND PRESENTING ELECTRONIC DOCUMENTS"
2	US 8,639,495	622▲ US 7,966,328	US 7,966,328	2▲ APP 20070208719: "SYSTEMS AND METHODS FOR ANALYZING SEMANTIC DOCUMENTS OVER A NETWORK"
3	US 8,639,497	608▲ APP 20070208719	7▲ US 6,466,928	4▲ APP 20110093449: "SEARCH ENGINE AND METHODOLOGY, PARTICULARLY APPLICABLE TO PATENT LITERATURE"
4	APP 20130173256	653▲ APP 20150310114	-1▼ APP 20070208719	2▲ APP 20050210008: "Systems and methods for analyzing documents over a network"
5	US 8,688,447	605▲ APP 20050210008	-1▼ APP 20150310114	29▲ APP 20160344770: "Automatic Phishing Email Detection Based on Natural Language Processing Techniques"

⚠️ Excerpt Alert

62.0%	APP 20120069131	▼ low ▼
62.0%	APP 20160086108	▼ low ▼
62.0%	APP 20140201126	▼ low ▼
61.0%	APP 20140075004	▼ low ▼

(%) Similarity of Idea as an excerpt of relevant document.

Key Phrases In NEIGHBORHOOD: small business innovation research phase (88.1), patent documents represent prior art (39.8), closely related patent documents (45.1), constructed term document matrix (43.3), approximately million patent documents (32.8), pca principal component analysis (31.5), small business innovation research (28.8), large document collections corpus (27.7), united states patent (26), relevant patent documents (25.5), national science foundation (25.4), search engine implemented (25.4), general purpose computing (25.1)