

Enriching Content with Semantic Tagging



ICIC 2013, Vienna

Molecular Connections, Bangalore, India

www.molecularconnections.com

Outline

- **Introduction to MC**
- **Content Enrichment – Concept**
- **Content Enrichment Use Case**
- **Key Take Aways**

About MC

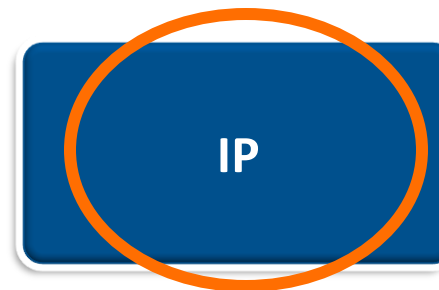
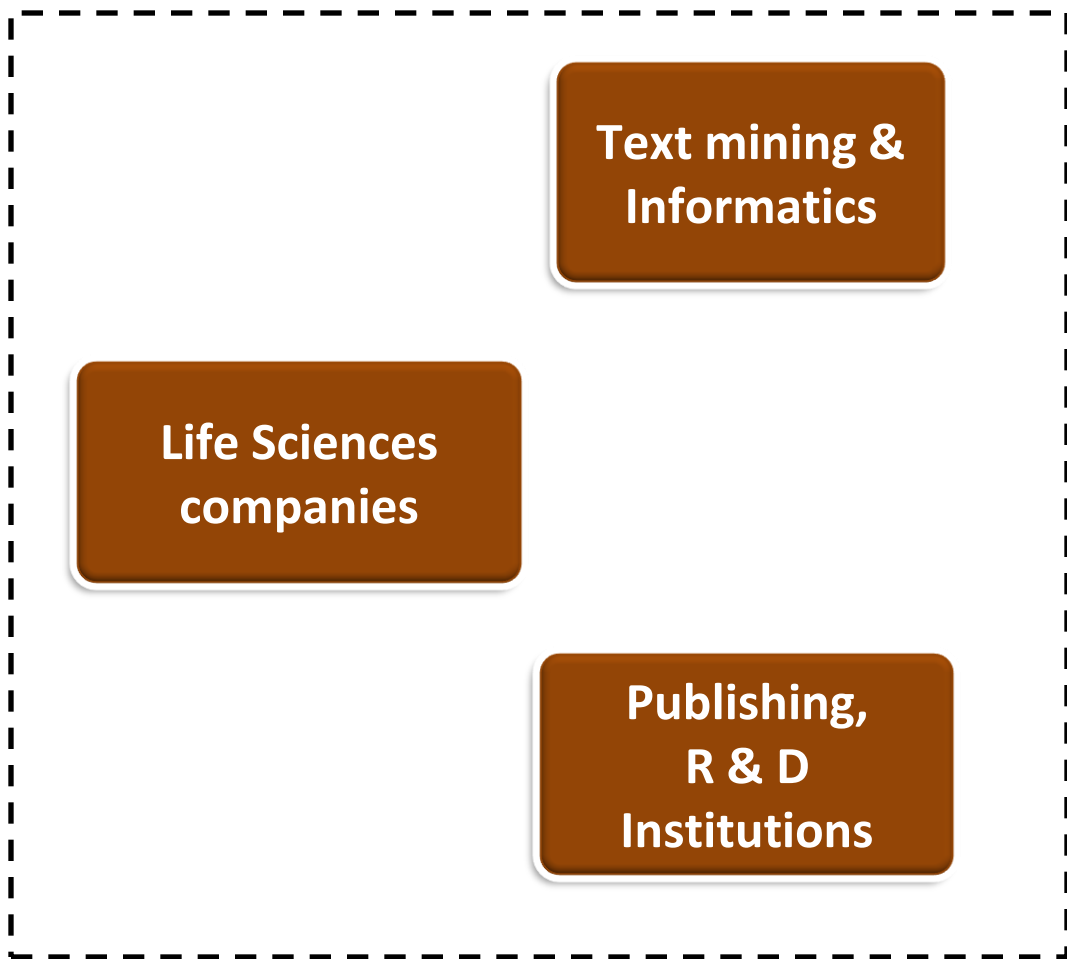
CORPORATE

- ❖ Established in 2001
- ❖ Executive team backed by renowned informaticans & strong advisory board -~ 1000 strong
- ❖ Scalable & state of the art infrastructure
- ❖ Global footprint
- ❖ Core Values: Customer focused, Quality, Ethics, Excellence, Accountability

OPERATIONS

- Information curation and annotation expertise
 - work with leading R & D Institutions , STM publishing & IP Search & Law Firms
- Right mix of human resources and scale
 - LifeScience (Bio – Chem), Engineering, IP, information and technology background
- Established workflow and processes to ensure quality and on time delivery
- ISO 27001: 2005 Certified knowledge management platforms and workflow systems

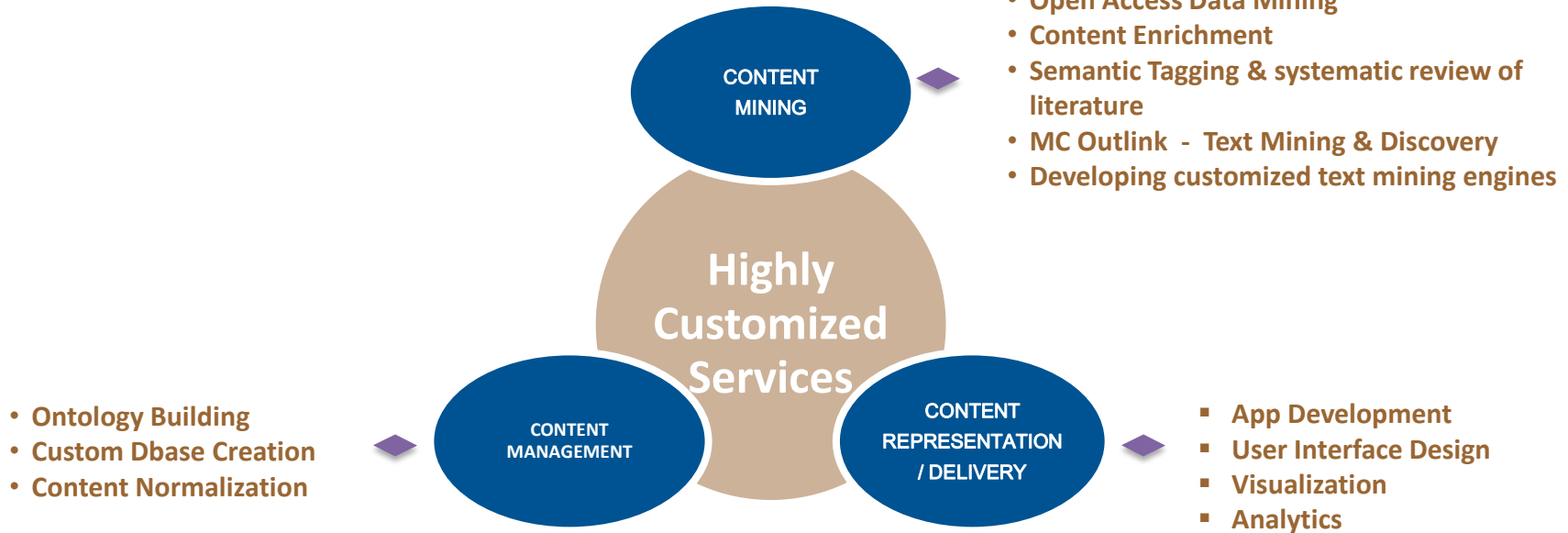
Verticals



- ✓ MCPaIRS
- ✓ MCDESIGN
- ✓ Patent Search Services

MC - Solutions

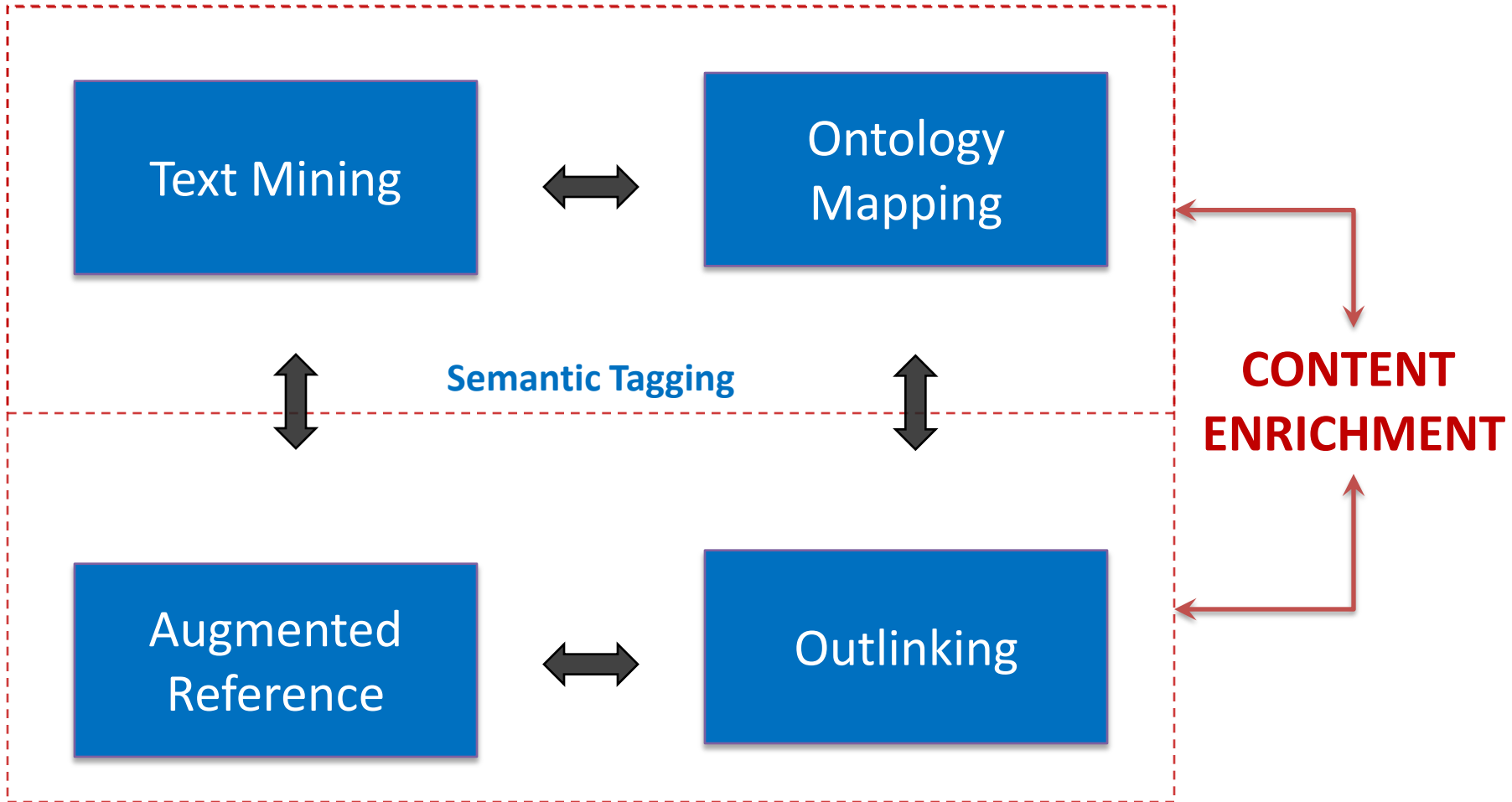
End ← to → End Solutions



Over 3500 Man Years of expertise



Enriching Content

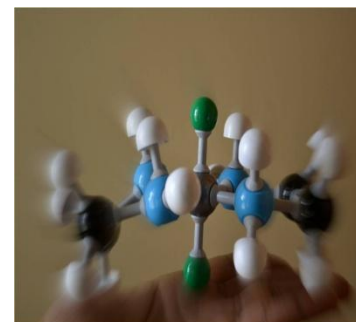
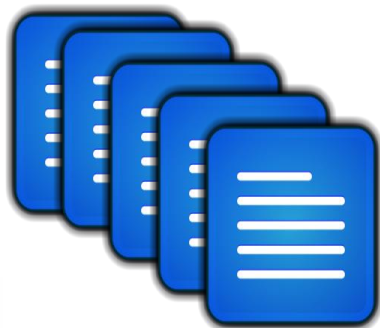


Why CE?

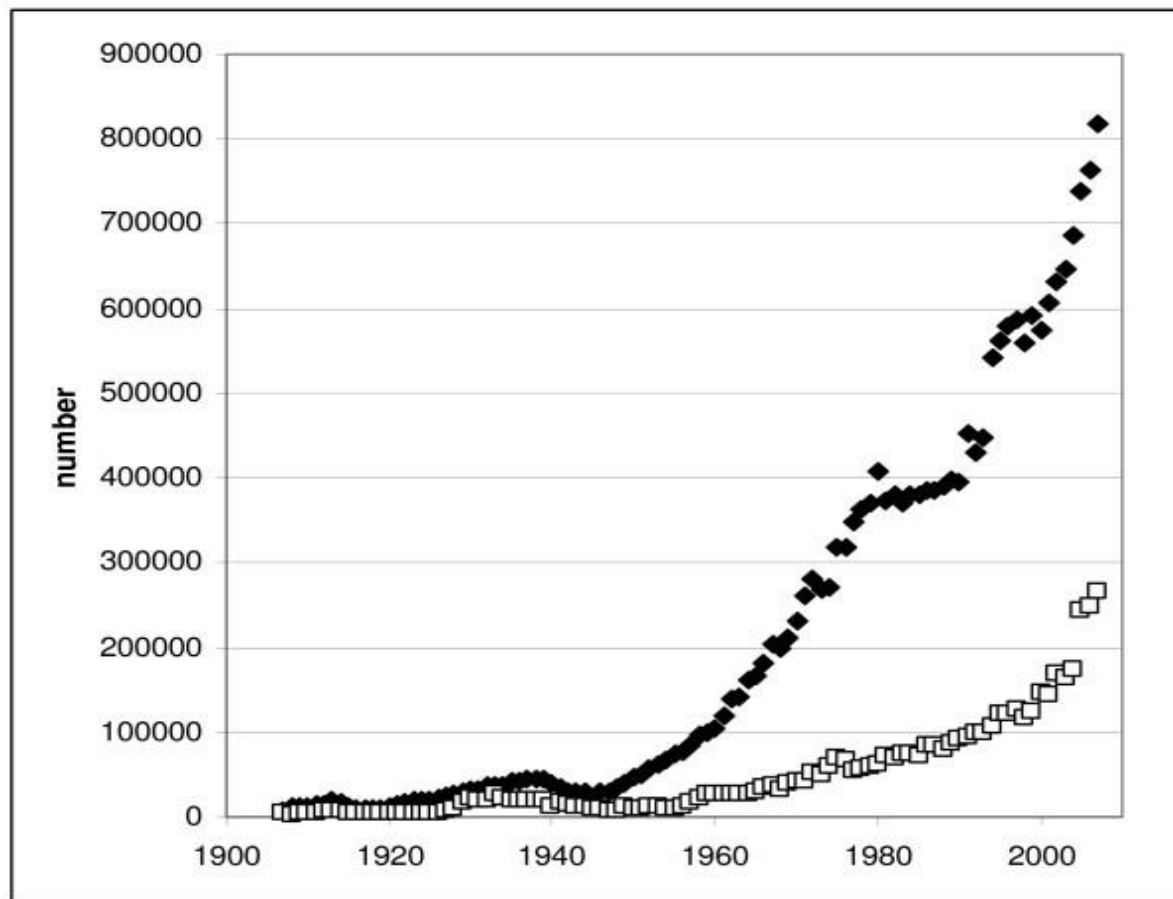
- Enables deeper knowledge discovery from diverse sources like patent, databases, journal etc.
- Semantic tagging ensures that different names of an entity are mapped to standard name and hence, searchable by any name.

For Instance: Discoverability is a challenge in pharma patents as entities of interest may be named differently in different patents by different authors.

- Publishers are quick to adopt CE, time to adopt it for patents?



Unlocking Small Data to Big Data



Number of articles (diamonds) and patents (open boxes) abstracted annually by *Chemical Abstracts Services*

Bachrach *Journal of Cheminformatics* 2009 1:2 doi:10.1186/1758-2946-1-2

DRUGBANK
Open Data Drug & Drug Target Database

UniProt

PubChem

ChEMBL

DBpedia

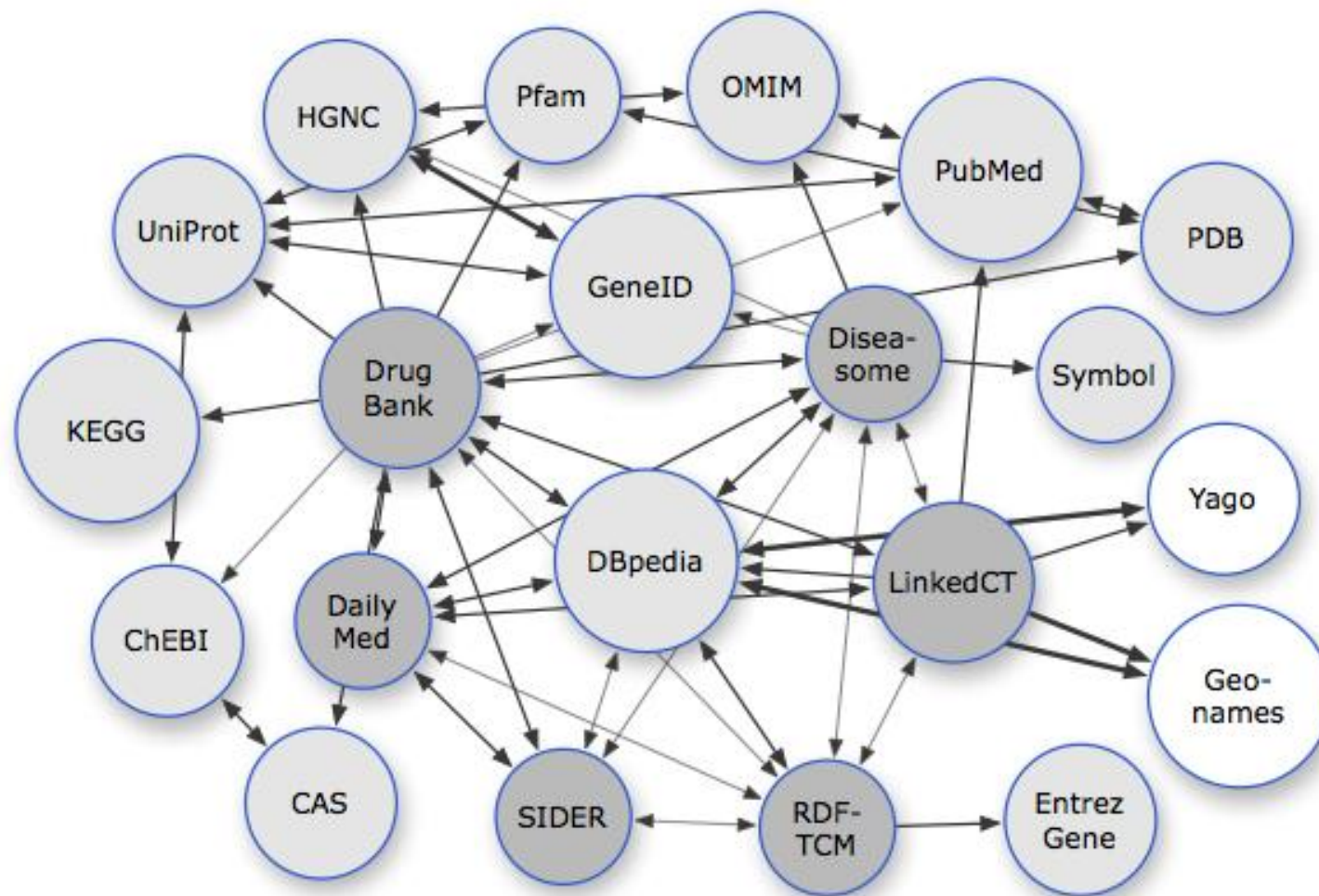
GO
the Gene Ontology

PubMed

e

Need Smarter Content

Leveraging Linked Data



Implementation - Content Enrichment Levels

- **Entity**

- Author/Assignee, Protein, Gene, Drug, Chemical, Disease, Reaction, Organism, Technology, Organization

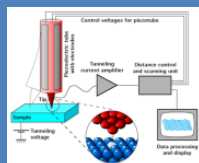
- **Document**

- Journal article
- Patent
- Book chapter



- **Others**

- Image
- Table
- Multimedia
- News links



Company	Division	Sector	Type
IBM_Combined_Company	IBM_Combined_Division	IBM_Combined_Sector	14025
Apple	IBM_Combined_Division	IBM_Combined_Sector	10121
Apple	Hardware	IBM_Combined_Sector	4900
Apple	Hardware	Business	100
Apple	Hardware	Consumer	3100
Apple	Software	IBM_Combined_Sector	5025
Apple	Software	Business	4900
Apple	Software	Consumer	475
Microsoft	IBM_Combined_Division	IBM_Combined_Sector	4900
Microsoft	Hardware	IBM_Combined_Sector	3000
Microsoft	Hardware	Business	800
Microsoft	Hardware	Consumer	3000
Microsoft	Software	IBM_Combined_Sector	2010
Microsoft	Software	Business	1010
Microsoft	Software	Consumer	1000



Content Enrichment – Use Case

MCPaIRS™ (Proprietary Indian Patent Database)

Molecular Connections Patent Information Retrieval System

·"Expertly , Manually Curated, Fully Searchable, Value Added Knowledgebase" of Full Text of Indian Granted and Applied Patents

·Caters to a diversified user-base of bench Scientists, Engineers, R&D Managers & Business Professionals.





WELCOME ginish.g@molecularconnections.com
[Change Password](#) [Logout](#)

[Home](#) | [About us](#) | [FAQ](#) | [Features](#) | [Quick Search](#) | [My MCPaIRS](#) | [Help](#) | [MCDESIGN](#)

Contact Us

Grant
 Application

[? Search for Title, Abstract, Applicant, Inventor etc.](#) [Quick Search](#)

About MCPaIRS



MCPaIRS is the first Indian Patent Full Text Database compiled by experts at Molecular Connections. The data is hand-curated by domain experts and provided in an easy to use xml format plus web interface. The database has full text of granted patents and published applications plus a well designed front page and legal status information for each document. The database is updated every week.

Customer support

Contact and Fax:
Phone : +91-80-41205016
Fax : +91-80-41208956

Email us:
ipservices@molecularconnections.com

Database last updated on
25th Sep, 2013

Links

[MCPaIRS - Subscription Model](#)

What is New?

Browse Indian Patents using your favorite gadgets!

- iPhone
- iPad
- Android App


for MCPaIRS is coming soon

 [About Us](#) | [Site Map](#) | [Privacy Policy](#) | [Contact Us](#) | [Copyright](#) | [Disclaimer](#)
© 2013 Molecular Connections (P) Ltd. All rights reserved.




MCPaIRS™ – Search

Home About us FAQ Features Quick search Help MCDESIGN Contact Us

 MCPaIRS
VALUE ADDED INDIAN PATENT DATABASE

"BICYCLIC PEPTIDOMIMETIC INHIBITORS" Grant Application

 Search for Title, Abstract, Applicant, Inventor etc.

Welcome ginish.g@molecularconnections.com

Filter By Filing Years

2011 (1)


Filter By Inventors

CASSONE, ANTONIO (1)
 DE BERNARDIS, FLAVIA (1)
 GARACI, ENRICO (1)
 GUARNA, ANTONIO (1)
 TRABOCCHI, ANDREA (1)

Filter By IPCCC

A61K 31/553 (1)
 C07D 498/08 (1)

ViewResults SnapView My MCPaIRS

Result of searching: ("BICYCLIC PEPTIDOMIMETIC INHIBITORS")  1

about **1 records**

Showing page 1 of 1 0

Select All

1 **Title:** **BICYCLIC PEPTIDOMIMETIC INHIBITORS OF ASPARTYL-PROTEASES FOR THE TREATMENT OF INFECTIOUS DISEASES**

Application Number: 4501/CHENP/2011

Applicants: UNIVERSITA' DEGLI STUDI DI FIRENZE

Inventors: CASSONE, ANTONIO ; DE BERNARDIS, FLAVIA ; GARACI, ENRICO ; TRABOCCHI, ANDREA ; GUARNA, ANTONIO

The present invention refers to 3-aza-bicyclo[3.2.1]octane derivatives of general formula (I) their preparation, use and pharmaceutical compositions useful in the treatment of pathologies associated with microbial pathogens expressing aspartyl- protease activity.

1

MCPaIRS™ – View Patent



Welcome ginish.g@molecularconnections.com

[Logout](#)

[Report a Issue](#)

(4501/CHENP/2011) **BICYCLIC PEPTIDOMIMETIC INHIBITORS** OF ASPARTYL-PROTEASES FOR THE TREATMENT OF INFECTIOUS DISEASES

Bibliographic Data | [Claims](#) | [Description](#) | [Drawings](#) | [Legal Status](#)

(21) **Application Number** : 4501/CHENP/2011
(22) **Filing Date** : 24/06/2011
(43) **Publication Date** : 14/09/2012
(71) **Applicant(s)** : UNIVERSITA' DEGLI STUDI DI FIRENZE;

(72) **Inventor(s)** : CASSONE, ANTONIO; DE BERNARDIS, FLAVIA; GARACI, ENRICO; TRABOCCHI, ANDREA; GUARNA, ANTONIO;

(51) **International Classifications** : C07D 498/08; A61K 31/553;
(30) **Priority** : EP 25/11/2008 08169901.9;
(86) **International Application Number and Date** : PCT/EP2009/065728 24/11/2009
(87) **International Publication Number** : WO/2010/060904A1

(54) **Title** : BICYCLIC PEPTIDOMIMETIC INHIBITORS OF ASPARTYL-PROTEASES FOR THE TREATMENT OF INFECTIOUS DISEASES

(57) **Abstract** :
The present invention refers to 3-aza-bicyclo[3.2.1]octane derivatives of general formula (I) their preparation, use and pharmaceutical compositions useful in the treatment of pathologies associated with microbial pathogens expressing aspartyl- protease activity.

Family : WO2010060904A1 AT540957T CN102224159A EP2189462A1
EP2189462B1 EP2189462B9 ES2380613T3 ES2380613T9
JP2012509863A US2011237577A1 US8445484B2 WO2010060904A1

Download



[About Us](#) | [Site Map](#) | [Privacy Policy](#) | [Contact Us](#) | © 2013 Molecular Connections (P) Ltd. All rights reserved.



Demo of actual full text document

Benefits of Semantic Search Cartridge Enabled MCPaIRS™

- ✓ All results in a single query
- ✓ Automatic Expansion of the query with all possible synonyms
- ✓ Broadening of the search query
- ✓ Complex search queries possible
- ✓ All the synonyms highlighted

Automatic Expansion of the query with all possible synonyms

Grant Application

Search for Title, Abstract, Applicant, Inventor etc.

[Logout](#)

0 Results

[SnapView](#)

[My MCPaIRS](#)

Result of searching: ("Diazepam-binding inhibitor-like 5" OR "Endozepine-like peptide" OR "Apozepam" OR "HCA88" OR "Tranimul" OR "ECI2_HUMAN" OR "Hepatocellular carcinoma-associated antigen 88" OR "diazepam" OR "ACBP_MOUSE" OR "DRS1" OR "Dodecenoyl-CoA isomerase" OR "Endozepine" OR "Dbil5" OR "DRS-1" OR "E-Pam" OR "Serenack" OR "ECI2" OR "Relanium" OR "Diastat acudial" OR "Diazepam" OR "Scriptopam" OR "DBIL5_RAT" OR "Stesolid" OR "Acyl-CoA-binding protein" OR "ACBP_RAT" OR "Delta(3),delta(2)-enoyl-CoA isomerase" OR "DBI" OR "PECI" OR "C16H13CIN20" OR "Peroxisomal 3,2-trans-enoyl-CoA isomerase" OR "Dizac" OR "Diazepam-binding inhibitor-related protein 1" OR "Diastat" OR "Dbi" OR "DBIL5_MOUSE" OR "Diazepam-binding inhibitor" OR "Q-pam" OR "Diazepam intensol" OR "Enoyl-CoA delta isomerase 2, mitochondrial" OR "ACBP_HUMAN" OR "Renal carcinoma antigen NY-REN-1" OR "Paxel")

about **2146 records**

Showing page 1 of 215 0

[EXPORT](#)

Select All

1 **Title:** **INDWELLING LUMINAL DEVICES**

Application Number: 916/KOLNP/2013

Applicants: W. L. GORE AND ASSOCIATES, INC.

Inventors: CULLY, EDWARD H. ; DUNCAN, JEFFREY B. ; MAULDING, MATTHEW E. ; TRAPP, BENJAMIN M. ; SCHONHOLZ, CLAUDIO

The invention comprises an indwelling medical device which is capable of delivering a therapeutic agent evenly along the length of the indwelling portion, including the outer wall, of the device.

2 **Title:** **METHODS AND NODES IN A WIRELESS COMMUNICATION NETWORK**

Application Number: 136/KOLNP/2013

Applicants: TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)

Automatic Expansion of the query with all possible synonyms

CD2 Grant Application

Search for Title, Abstract, Applicant, Inventor etc.

Result of searching: ("CIN85" OR "T-cell surface antigen T11/Leu-5" OR "Human Src family kinase-binding protein 1" OR "Mets1" OR "CS1" OR "SLAM family member 9" OR "Mesenchyme-to-epithelium transition protein with SH3 domains 1" OR "CD2-associated protein" OR "SLAM family member 7" OR "CD2AP" OR "PPIP1_HUMAN" OR "CD2F10" OR "H-PIP" OR "Cas ligand with multiple SH3 domains" OR "KIAA1178" OR "CD2 family member 10" OR "CD2 subset 1" OR "CD2-binding protein 1" OR "CD2-binding protein 3" OR "Membrane protein FOAP-12" OR "CD2_RAT" OR "SRBC" OR "SLAF7_HUMAN" OR "Cd2" OR "Lymphocyte antigen 37" OR "Adapter protein CMS" OR "OX-34 antigen" OR "SH3KBP1" OR "Ly-37" OR "LFA-2" OR "CD2_HUMAN" OR "SH3K1_HUMAN" OR "Proline-serine-threonine phosphatase-interacting protein 1" OR "SH3 domain-containing kinase-binding protein 1" OR "Cd2ap" OR "CD2B2_MOUSE" OR "CD2 antigen cytoplasmic tail-binding protein 2" OR "CD84 homolog 1" OR "CD2BP1" OR "Novel Ly9" OR "CD2BP2" OR "CD2" OR "PSTPIP1" OR "Rosette receptor" OR "Cbl-interacting protein of 85 kDa" OR "T-cell surface antigen CD2" OR "CD2AP_MOUSE" OR "SLAMF9" OR "CD2-like receptor-activating cytotoxic cells" OR "CD2B2_HUMAN" OR "U5 snRNP 52K protein" OR "SLAMF7" OR "CD2AP_HUMAN" OR "LFA-3 receptor" OR "Cd2bp2" OR "Protein 19A" OR "Erythrocyte receptor" OR "SLAF9_HUMAN" OR "CD2_MOUSE")

about **1997 records**

Showing page 1 of 200

MC DESIGN 0

Select All

1 **Title:** 5-(BIPHENYL-4-YL)-3-PHENYL-1,2,4-OXADIAZOLYL DERIVATIVES AS LIGANDS ON THE SPHINGOSINE 1-PHOSPHATE (S1P) RECEPTORS

Application Number: 320/KOLNP/2013

Applicants: MERCK SERONO S.A.

Inventors: BOMBRUN AGNES ; QUATTROPANI ANNA ; GONZALEZ JEROME ; DORBAIS JEROME ; KNIGHT CHRIS ; BAKER-GLENN CHARLES

The present invention provides compounds of Formula (I), as selective S1 P1 inhibitors, as well as their use for treating multiple sclerosis and other diseases.



2 **Title:** ANTI-CD48 ANTIBODIES AND USES THEREOF

Application Number: 832/KOLNP/2013

Applicants: DEGENERON BIOMEDICALS INC

Multiple key-words highlighted for the search: VEGF

Select All



Title: DUAL INHIBITORS OF MET AND VEGF FOR THE TREATMENT OF CASTRATION- RESISTANT PROSTATE CANCER AND OSTEOBLASTIC BONE METASTASES

Application Number: 998/KOLNP/2013

Order now

Applicants: EXELIXIS, INC.

Inventors: SMITH, DAVID ; HUSSAIN, MAHA

 Legal Status

This invention is directed to the treatment of cancer, particularly castration-resistant prostate cancer and osteoblastic bone metastases, with a dual inhibitor of MET and VEGF.



Title: ROCK2 AND ROCK3 TWO NEW GAIN-OF-FUNCTION VARIANTS OF THE CYTOKININ RECEPTORS AHK2 AND AHK3

Application Number: 424/CHENP/2012

Order now

Applicants: FREIE UNIVERSITAT BERLIN

Inventors: SCHMULLING, THOMAS ; WERNER, TOMAS ; BARTRINA Y MANNS, ISABEL ; BRAUN, HELEN

 Legal Status

The present invention relates to two new gain of function variants of the cytokinin receptor proteins AHK2 and AHK3 namely rock2 and rock3 to transgenic organisms comprising at least one of said new gain-of-function cytokinin receptor variants and to a method for the manufacturing of a transgenic plant comprising at least one of the new gain-of-function variants.




Title: BIOLOGICAL MARKERS FOR MONITORING PATIENT RESPONSE TO VEGF ANTAGONISTS

Application Number: 58/MUMNP/2012

Order now

Applicants: F. HOFFMANN-LA ROCHE AG (CH)

Inventors: BAIS, CARLOS ; BRAUER, MATTHEW ; KAMINKER, JOSHUA ; SINGH, MALLIKA

 Legal Status

The invention provides methods and compositions to detect expression of one or more biomarkers for monitoring the effectiveness of treatment of with VEGF antagonists. The invention also provides methods for identifying and treating patients who are likely to be responsive to VEGF antagonist therapy. The invention

Complex Queries can be performed by using operators

Salbutamol AND CD48 Grant Application

Search for Title, Abstract, Applicant, Inventor etc.

[Logout](#)

[View Results](#)

[SnapView](#)

[My MCPaIRS](#)

Result of searching: ("Volmax" OR "C13H21NO3" OR "SALBUTAMOL" OR "Ventolin HFA" OR "Albuterol" OR "Proventil" OR "Combivent" OR "Duoneb" OR "Salbutamol" OR "ALBUTEROL" OR "Proair HFA" OR "PROVENTIL-HFA" OR "Proventil HFA" OR "Accuneb" OR "Ventolin") AND ("sgp-60" OR "Signaling lymphocytic activation molecule 2" OR "TCT.1" OR "CD48_RAT" OR "CD48 antigen" OR "BCM1 surface antigen" OR "Leukocyte antigen MEM-102" OR "CD48_MOUSE" OR "BLAST-1" OR "MRC OX-45 surface antigen" OR "Cd48" OR "CD48" OR "BLAST1" OR "SLAM family member 2" OR "HM48-1" OR "B-lymphocyte activation marker BLAST-1" OR "BCM1" OR "Bcm-1" OR "CD48_HUMAN")

about **4 records**

[EXPORT](#)

Select All

1

Title: NICOTINE IMMUNONANTHERAPEUTICS

Application Number: 3540/DELNP/2011

[Order now](#)

Applicants: MASSACHUSETTS INSTITUTE OF TECHNOLOGY ; PRESIDENT AND FELLOWS OF HARVARD COLLEGE ; THE BRIGHAM AND WOMEN'S HOSPITAL, INC.

Inventors: MATTEO, IANNAcone ; ULRICH, VON ANDRIAN ; OMID, C. FAROKHZAD ; FRANK, ALEXIS ; PAMELA, BASTO ; JINJUN, SHI ; ELLIOTT, ASHLEY MOSEM ; ROBERT, LANGER ; ELENA, TONTI

[Legal Status](#)

The present invention pro-vides compositions and systems for delivery of nanocarriers to cells of the immune sys-tem. The invention provides nanocarriers capable of stimulating an immune response in T cells and/or in B cells. The invention provides nanocarriers that comprise an im-munofeature surface having a plurality of nicotine moieties. The invention provides pharmaceutical compositions comprising nanocarriers. The present invention pro-vides methods of designing, manufacturing, and using nanocarriers and pharmaceutical compositions thereof For example, the present invention describes nanocarriers capable of eliciting an immune response and the production of anti-nicotine antibodies.

Boolean search is performed

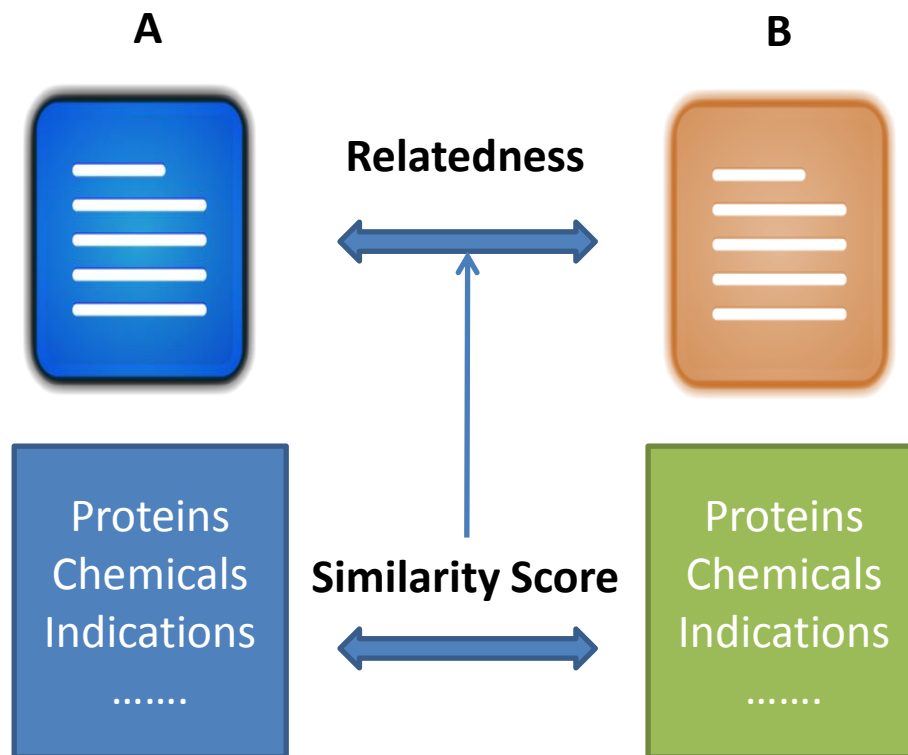
2

Title: TARGETING OF ANTIGEN PRESENTING CELLS WITH IMMUNONANTHERAPEUTICS

Sample queries with Semantic Search Cartridge

No	Query	No of results in iPairs	No of results in mcpairs	No of results in mcpairs with semantic search cartridge
1	Salbutamol	27	1560	2548
2	Amethocaine	0	58	954
3	Diazepam	4	1725	2146
4	Valsartan	84	1372	1429
5	Imatinib	65	1703	1999
6	Tamoxifen	16	3950	4190
7	Aspirin	61	5679	6427
8	Paracetamol	74	1161	3696
9	MyoD	2	130	138
10	Pax3	1	49	56
11	Sox9	0	39	58
12	FGF10	0	43	131
13	VEGF	192	4808	6058
14	BMP2	5	137	214
15	Salbutamol AND CD48	0	0	4

Benefit - Identifying Related Patents



Content Enrichment Approaches

- **Manual**
 - ✓ high quality, costly, not scalable, slow
- **Automated**
 - ✓ fast, quality below par, cost effective, scalable
- **Hybrid**
 - ✓ high quality, cost effective, scalable, reasonable speed

Molecular Connections is a pioneer in the use of hybrid approach to content enrichment

Key Takeaways

- ✓ **Content Enrichment can improve search and retrieval immensely**
- ✓ **?? CE can be looked at various levels**
 - **Biology / chemistry / both / authors etc.**
- ✓ **You can bring the Web into the document through CE**
 - **e.g. Augmented reference cards**
- ✓ **Growing Adoption of Content Enrichment**
 - **Publishing (Early adopters)**
 - **Patents**



Thank You

Molecular Connections

www.molecularconnections.com