

Visualizaciones bibliométricas con indicadores dependientes e independientes del tamaño utilizando la metodología ViBlioSOM

Bibliometric visualizations with size-dependent and size-independent indicators using the ViBlioSOM methodology

Anette Lozano Díaz¹, José Luis Jiménez-Andrade², Elio Atenógenes Villaseñor-García³ & Humberto Andrés Carrillo-Calvet⁴

¹Estudiante de Posgrado en Bibliotecología y Estudios de la Información (UNAM), México, ibis.alozano@gmail.com

²Laboratorio de Dinámica no Lineal, Departamento de Matemáticas, Facultad de Ciencias, UNAM y Centro de Ciencias de la Complejidad C3 (UNAM), Laboratorio de Dinámica no Lineal, Departamento de Matemáticas, Facultad de Ciencias, UNAM y Centro de Ciencias de la Complejidad C3, UNAM, chilti@gmail.com

³Centro de Investigación e Innovación en Tecnologías de la Información, México, atenogenes@gmail.com

⁴Laboratorio de Dinámica no Lineal, Departamento de Matemáticas, Facultad de Ciencias y Centro de Ciencias de la Complejidad C3 (UNAM), México, hcarrillocalvet@gmail.com

Resumen

Las metodologías multidimensionales permiten realizar estudios bibliométricos que combinan, de manera no lineal, varios aspectos de la actividad científica. De esta manera, las comparaciones que se realizan entre unidades científicas (e.i. instituciones, países), pueden considerar, simultáneamente, diferentes facetas de la actividad científica, como productividad, impacto, eficiencia y eficacia. Sin embargo, al igual que en los rankings

unidimensionales, es necesario tener en cuenta la naturaleza de los indicadores utilizados para realizar las interpretaciones adecuadamente. El objetivo general de este trabajo es analizar el efecto que tiene el uso de indicadores dependientes e independientes de tamaño en la metodología ViBlioSOM; técnica multiparamétrica y de visualización basada en distancia. Para ilustrarlo, nos apoyamos en un estudio multiparamétrico que compara el perfil científico de un grupo de países con productividad similar a México. ViBlioSOM permitió identificar automáticamente el perfil científico en cuanto a la eficiencia, impacto y excelencia para 68 países; así como, determinar los valores atípicos que se destacan como perfiles peculiares. No obstante, los resultados con indicadores independientes y dependientes de tamaño en ViBlioSOM ofrecieron perspectivas diferentes.

Palabras clave: Indicadores dependientes de tamaño; indicadores independientes del tamaño; Redes neuronales; Mapas auto-organizados; minería de datos cuantitativa; ViBlioSOM

Abstract

Multidimensional methodologies allow bibliometric studies that combines, nonlinearly, several aspects of scientific activity. Thus, the comparisons made between scientific units (i.e. institutions, countries) may consider, simultaneously, different facets of scientific activity, such as productivity, impact, efficiency and effectiveness. However, as in the one-dimensional rankings, it is necessary to take into account the nature of indicators in order to perform interpretations properly. The overall objective of this work is to analyze the effect of the use of size-independent and size-dependent indicators on the ViBlioSOM methodology; multiparametric and distance-based visualization technique. To illustrate this, we rely on a multiparametric study that compares the scientific profile of a group of countries with similar productivity to Mexico. ViBlioSOM allowed to automatically identify the scientific profile in terms of efficiency, impact and excellence for 68 countries; as well as determine outliers that stand out with distinctive profiles. Nevertheless, the results using size-independent and size-dependent on ViBlioSOM, offered different perspectives.

Keywords: size dependent indicators; size independent indicators; neural networks; self-organizing maps; data mining scientiometric; ViBlioSOM

Introducción

Las metodologías multiparamétricas o multidimensionales permiten realizar estudios bibliométricos que combinan, de manera no lineal, varios aspectos de la actividad científica, permitiendo contrastar productividad, impacto, eficiencia y excelencia [multiparametric]. Aunque estas técnicas surgieron para realizar mapeos de la ciencia, es decir, visualizaciones bibliométricas, son herramientas que pueden ser utilizadas para construir ordenamientos bidimensionales de unidades científicas con menor pérdida de información que los rankeos unidimensionales (Van Eck, N. J., & Waltman, L. 2010, Guzmán-Sánchez, 2009).

Sin entrar a la discusión sobre la validez de los indicadores, o la imposibilidad de incluir en los indicadores los recursos invertidos para medir eficiencia (Bornmann, L., & Haunschild, R. 2016), a la pregunta, ¿qué deberíamos medir? (Glänzel, W., Thijs, B., & Debackere, K. 2016) creemos que la respuesta es, lo que nos permita valorar la mayor cantidad de aspectos. No obstante, también creemos que apoyarse en herramientas y técnicas para visualizar y analizar la información bibliométrica es tan importante como la integración de indicadores de diferente naturaleza.

ViblioSOM es una metodología basada en el algoritmo SOM (red neuronal artificial) (Kohonen, 1990) para construir mapas de conocimiento. En primera instancia, podemos usar la metodología para construir mapas de la ciencia haciendo análisis de coocurrencia (Guzmán-Sánchez, 2009), pero si usamos indicadores de actividad e impacto, es posible realizar estudios bibliométricos multiparamétricos (Villaseñor-García, 2016b).

Al igual que en los rankeos unidimensionales, los mapas obtenidos con ViBlioSOM dependen de los indicadores de entrada. En este trabajo nos apoyamos en un caso de estudio

para mostrar el efecto que tiene el uso de indicadores dependientes de tamaño (Abramo, G., & D'Angelo, C. A. 2016) sobre las técnicas de visualización basadas en distancias; particularmente sobre la metodología ViBlioSOM.

Objetivo General

El objetivo general de este estudio es analizar la aplicación de indicadores de tamaño dependiente e independiente para determinar el perfil cientométrico de un grupo de países con similar producción científica que México, con el empleo de la metodología ViBlioSOM.

Fuentes de información

La fuente principal de información fue el *EssentialScienceIndicators*(ESI). El ESI es un recurso de análisis que forma parte de la plataforma InCites, de Thomson Reuters. La extracción de los datos de producción e impacto se realizó con la interfaz web del ESI, en la pestaña *Indicators*, filtrando por país. El promedio de citas por campo de investigación se obtuvo en la pestaña *Field Baselines*. La información se descargó en junio de 2015, para el periodo 2005-2015.

Para normalizar los indicadores, se utilizó el número de habitantes de cada país. La información fue extraída del banco de datos del Banco Mundial.

Indicadores

El análisis multiparamétrico se realizó considerando indicadores que describen tres aspectos fundamentales de la ciencia: desarrollo científico, impacto y excelencia científica.

En este trabajo se usaron los indicadores *HighlyCitedPaper* y *Hot Paper* proporcionados por ESI, como indicadores de excelencia científica. En la primera parte del estudio se utilizaron tal y como fueron descargados. Para la segunda parte, se dividieron por el número de documentos del periodo para hacerlos independientes de tamaño.

Highly cited papers (HCP): número de artículos que recibieron un total de citas superior al 1% de citas cuando se compara con todos los artículos publicados en el mismo año y en la misma disciplina.

Hot papers (HOT): número de artículos en los dos últimos años que recibieron una cantidad de citas suficientes en los dos últimos meses para ubicarse en el 0,1% superior en comparación con otros artículos pares del mismo período y disciplina.

Impacto Normalizado (NI). Para calcular el impacto normalizado nos auxiliamos de las 22 áreas científicas definidas en el ESI. De esta forma fue posible calcular, para cada país, el promedio de citas recibidas por documento, corregido por el promedio de citas que recibió cada una de las 22 áreas en cada año. Este indicador es independiente de tamaño.

Índice de Desarrollo Científico (Scientific Development Index - **SDI**) [ref tesis Ibis]. Definido como la proporción de documentos por habitante.

Metodología

La Figura 1. ilustra los pasos de la metodología ViBlioSOM que se aplicaron en esta investigación.

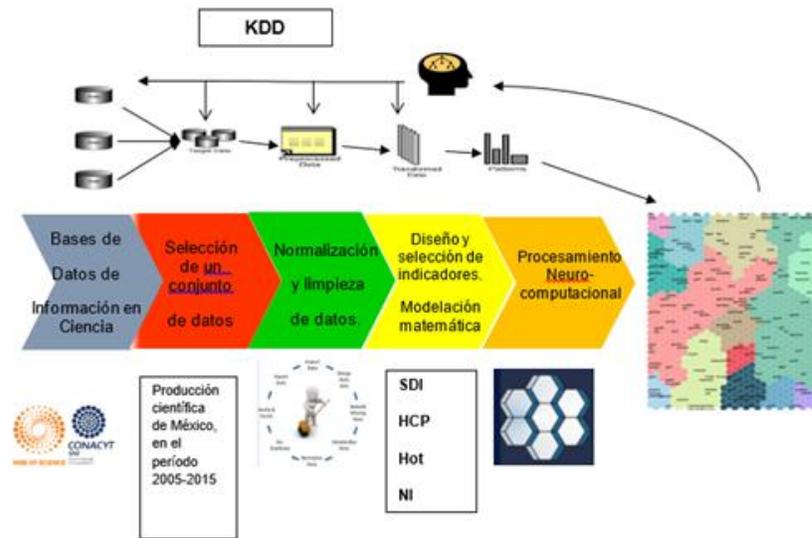


Figura 1. Metodología ViBlioSOM: etapas del proceso de descubrimiento, análisis y visualización de información cuantitativa

La siguiente tabla describe los pasos realizados para aplicar la metodología ViBlioSOM (Guzmán-Sánchez, 2009; Villaseñor-García, 2016a).

Tabla 1. Guía para aplicar la metodología ViBlioSOM

<p>1. En un primera instancia se extrajeron los datos desde EssentialScienceIndicators y la información exhibida por el Banco Mundial de los habitantes por país.</p>
<p>2. Se conformó una tabla con los indicadores como columnas y los países como renglones. En el ámbito del ViBlioSOM, la tabla es una matriz y los renglones son vectores. Con el propósito de que los indicadores tengan el mismo peso durante la identificación de cúmulos que hace el SOM, es necesario ponderarlos, es decir, equilibrar sus rangos de variabilidad. Este preprocesamiento consiste en dividir cada indicador (toda la columna) por el mayor de los indicadores en dicha columna. Después de esta operación, todos los indicadores tienen un rango de variación entre cero y uno (considerando que todos los indicadores son positivos o igual a cero).</p>
<p>3. Una vez que se tiene la matriz, se utiliza la herramienta de software para correr el algoritmo SOM (LabSOM, 2015). Los parámetros que se configuran son el tamaño del mapa (lado x ancho) y el número de iteraciones, el resto de los parámetros del algoritmo se conservan en su valor por defecto.</p>

4. Al término del entrenamiento, se analizan dos tipos de visualizaciones: mapas de componentes y mapas de clustering (Villaseñor-García, 2016a). Cada componente es un indicador. Por ello, se obtiene un mapa de componente por cada indicador. La interpretación de estos mapas permite analizar posibles correlaciones entre los indicadores. También son útiles para visualizar la distribución del indicador e identificar visualmente aquellos países que tienen un valor alto o bajo en el indicador. En ambas visualizaciones se mapean las etiquetas que identifican a los países. En los mapas de componentes, la escala de colores que se usa va de verde a rojo, pasando por el amarillo y el naranja. Las zonas rojas corresponden a valores altos del indicador y las zonas verdes a zonas con valores bajos.

Resultados y Discusión

Para ilustrar el tipo de conclusiones que se obtienen cuando se usan indicadores dependientes de tamaño, presentamos un análisis multiparamétrico para comparar el perfil de desempeño científico de países con similar índice de desarrollo científico a México.

El ranking descargado desde el ESI, documentos en el *web of science* y las citas, cuenta con 156 países. México ocupa el lugar 31 al ordenar por número de documentos y el lugar 37 al ordenar por número de citas. En ambos ordenamientos, la lista es encabezada por Estados Unidos de Norteamérica, Alemania, Inglaterra, China, Francia y Japón. Al ordenar por citas/artículos, México pasa al lugar 103, y los países que encabezan la lista son Gambia, Bermudas, Islandia, Panamá, Gibraltar y Suiza (en ese orden), desplazando a Estados Unidos de Norteamérica al lugar 12, a Inglaterra al 14 y a Alemania al lugar número 20. Quizás el caso más dramático sea Bermudas, que pasó del lugar 142 en el ordenamiento por citas, al segundo lugar en el ordenamiento de citas por artículo. Sin embargo, si ordenamos por impacto normalizado, que como sabemos es un indicador que toma en cuenta los estilos de citación de las áreas científicas, Bermudas pasa al lugar 145. Estos movimientos también se observan en las visualizaciones basadas en distancias; particularmente, en los mapas que se obtienen con la metodología ViBlioSOM, encontramos distribuciones totalmente distintas, que nos llevan a conclusiones diferentes.

Para el ejemplo consideramos 68 países que, de acuerdo al ranking SDI, se encuentran alrededor de México (34 arriba y 34 por debajo). De este grupo de países se excluyó a China porque, este país presenta altos valores para los cuatro indicadores estudiados (SDI= 1.21; HCP= 15863; Hot= 574; NI= 0.74), lo cual la convierte en un outlier, que no deja observar el comportamiento del resto de los países. Se realizaron dos experimentos, utilizando cuatro indicadores: SDI, HCP, Hot y NI. Tanto el SDI como el NI son indicadores independientes de tamaño. En el primer experimento usamos HCP y Hot sin normalizar (i.e. sin convertirlo en un indicador independiente de tamaño) y en el segundo se normalizaron. En la Figura 2 se muestran los mapas del primer experimento. En estos mapas queremos destacar el grupo de países ubicado en la esquina superior izquierda, compuesto por Portugal (PRT), Iran (IRN), Sudáfrica (ZAF), Turquía (TUR) y República Checa (CZE), cuya producción de excelencia es alta (altos valores en HCP y Hot) y su desarrollo científico es bajo (incluso Portugal y República Checa tienen un nivel por debajo de la media). La conclusión que se hace es que, este grupo de países, al igual que otros presentes en el mapa, tienen un perfil científico similar.

En el segundo experimento los cuatro indicadores son independientes de tamaño. Vemos que los cinco países se han separado; Portugal y República Checa permanecieron juntos, con Sudáfrica en el mismo cluster, pero ligeramente separado. Turquía e Irán pasaron a otro cluster. Esto quiere decir que no tienen un perfil científico similar. Así como este, se observan varios casos de separaciones y agrupaciones.

Queremos destacar también los países que, teniendo un perfil científico único, ocupan en la figura 2 su propio cluster. Se trata de Nueva Zelanda (NZL), Gales (WAL) e Islandia (ISL). Los dos primeros, según los resultados del segundo experimento, realmente no tienen un perfil científico tan diferenciado y se agruparon con Irlanda del Norte y Luxemburgo. Por su parte Islandia permaneció en su propio cluster. Ya en los primeros mapas destacaba por ser el país con mayor desarrollo científico e impacto, pero no así por los indicadores de excelencia (HCP y HOT). Sin embargo, en la Figura 3, vemos que tiene un valor alto en HCP

y un valor por arriba de la media en HOT. Georgia (GEO) es un caso que valdría la pena investigar, pues, aunque tiene bajo impacto y desarrollo científico, tiene alta excelencia científica, lo que lo convierte en un caso único.

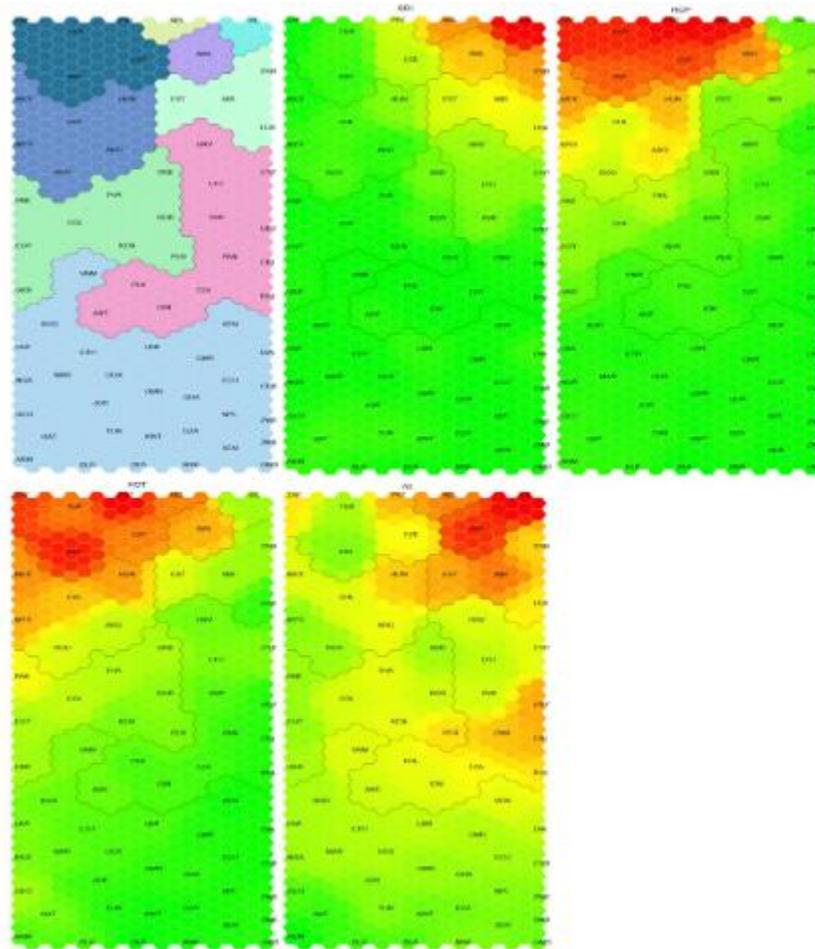


Figura 2. Cluster identificado con la técnica SelfOrganizedMaps para 68 países que tienen perfiles similares a México, HCP y Hot como indicadores de dependiente de tamaño

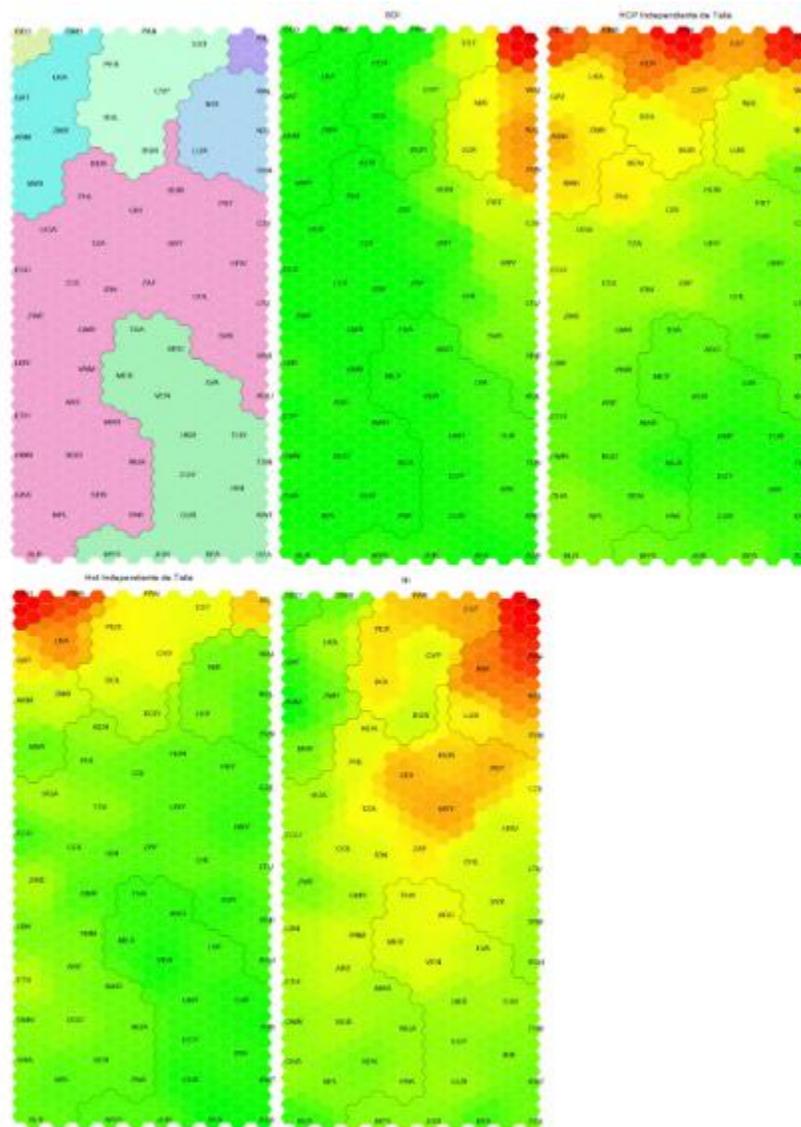


Figura 3. Cluster identificado con la técnica SelfOrganizedMaps para 68 países que tienen perfiles similares a México, HCP y Hot como indicadores independientes de tamaño

Conclusiones

La metodología ViBlioSOM es útil para llevar a cabo análisis multifactoriales y caracterizaciones cuantitativas. En el trabajo presentado se analizó un grupo de países que tienen un índice de desarrollo científico similar a México. ViBlioSOM nos permitió identificar automáticamente el perfil científico en cuanto a la eficiencia, impacto y excelencia

para 68 países; así como, determinar los valores atípicos que se destacan con perfiles peculiares. Este análisis nos permitió mostrar cómo la elección de indicadores dependientes de tamaño produce una caracterización multidimensional del perfil científico, distinta a la que se obtiene al usar indicadores independientes de tamaño. Las diferencias observadas no implican que alguna de las dos caracterizaciones sea errónea, simplemente se trata de dos perspectivas que pueden ser aprovechadas en escenarios distintos.

Agradecimientos

La autora Ibis A. Lozano Díaz (No. CVU 637915) agradece al Consejo Nacional de Ciencia y Tecnología de México (CONACYT) y el posgrado de Bibliotecología y Estudios de la Información de la UNAM por el apoyo para lograr el presente estudio, el cual es parte de la investigación de tesis para obtener el grado de Master.

Referencias

Abramo, G., & D'Angelo, C. A. (2016). A farewell to the MNCS and like size-independent indicators. *Journal of Informetrics*, 10(2), 646-651.

Bornmann, L., & Haunschild, R. (2016). Efficiency of research performance and the glass researcher. *arXiv preprint arXiv:1605.09515*.

Glänzel, W., & Debackere, K. (2009). On the "multi-dimensionality" of ranking and the role of bibliometrics in university assessment.

Glänzel, W., Thijs, B., & Debackere, K. (2016). Productivity, performance, efficiency, impact-What do we measure anyway?. Some comments on the paper "A farewell to the MNCS and like size-independent indicators" by Abramo and D'Angelo. *Journal of Informetrics*.

Guzmán-Sánchez, M. V. (2009). ViBlioSOM: Metodología para la visualización de información métrica con mapas auto-organizados. Universidad de La Habana, La Habana, Cuba.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464-1480.

LabSOM, 2015. Laboratorio para el Análisis del SOM.
<http://www.dynamics.unam.edu/DinamicaNoLineal3/labsom.htm>

Thelwall, M. (2016). Not dead, just resting: The practical value of per publication citation indicators. *arXiv preprint arXiv:1606.00193*.

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538.

Villaseñor-García, E.-A. (2016a). Tesis de Doctorado en revisión.

Villaseñor-García, E.A., Arencibia Jorge, R., & Carrillo-Calvet, H. (2016b). Multiparametric characterization of scientometric performance profiles assisted by neural networks: a study of Mexican higher education institutions. Por aparecer en *Scientometrics*.

ANEXO. Abreviaturas de países

PAÍSES	SIGLAS
ALGERIA	DZA
ARGENTINA	ARG
ARMENIA	ARM
BANGLADESH	BGD
BELARUS	BLR
BOLIVIA	BOL
BULGARIA	BGR
BURKINA FASO	BFA
CAMEROON	CMR
CHILE	CHL
COLOMBIA	COL
COSTA RICA	CRI
CROATIA	HRV
CUBA	CUB
CYPRUS	CYP

CZECH REPUBLIC	CZE
ECUADOR	ECU
EGYPT	EGY
ESTONIA	EST
ETHIOPIA	ETH
GAMBIA	GMB
GHANA	GHA
HUNGARY	HUN
ICELAND	ISL
INDONESIA	IDN
IRAN	IRN
JORDAN	JOR
KENYA	KEN
KUWAIT	KWT
LATVIA	LVA
LEBANON	LBN
LITHUANIA	LTU
LUXEMBOURG	LUX
MALAWI	MWI
MALAYSIA	MYS
MEXICO	MEX
MOROCCO	MAR
NEPAL	NPL
NEW ZEALAND	NZL
NIGERIA	NGA
NORTHERN IRELAND	NIR
OMAN	OMN
PAKISTAN	PAK

PANAMA	PAN
PERU	PER
PHILIPPINES	PHL
PORTUGAL	PRT
QATAR	QAT
REPUBLIC OF GEORGIA	GEO
ROMANIA	ROU
SENEGAL	SEN
SERBIA	SRB
SLOVAKIA	SVK
SLOVENIA	SVN
SOUTH AFRICA	ZAF
SRI LANKA	LKA
TANZANIA	TZA
THAILAND	THA
TUNISIA	TUN
TURKEY	TUR
UGANDA	UGA
UKRAINE	UKR
UNITED ARAB EMIRATES	ARE
URUGUAY	URY
VENEZUELA	VEN
VIETNAM	VNM
WALES	WAL
ZAMBIA	ZMB
ZIMBABWE	ZWE