

정보자원과 코퍼스

배 광 희

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《새로운 과학기술분야를 개척하기 위한 사업도 전망성있게 밀고나가야 합니다. 나라의 과학기술을 세계적수준에 올려세우자면 발전된 과학기술을 받아들이는것과 함께 새로운 과학기술분야를 개척하고 그 성과를 인민경제에 적극 받아들여야 합니다.》(《김정일선집》 증보판 제11권 138~139페이지)

지난 시기 언어연구는 주로 언어에 대한 교육과 사용을 위해 언어자체에 대한 범위에 한정되어 진행되어왔다고 볼수 있다.

그러나 오늘날 언어연구는 지식경제시대의 요구에 맞게 나라의 경제를 발전시키고 인민생활을 향상시키기 위한 측면에서도 연구되어야 한다.

지금까지 자원은 일정한 목적에 쓰이는 인적 또는 물질적원천으로만 보았다. 이러한 인적 또는 물질적자원은 언제나 경제를 발전시키고 인민생활을 높이는데 동원될수 있는 원천으로, 내부예비로 된다.

그러나 지식경제시대에 와서 자원에 대한 정의에서는 변화가 일어나고있다. 즉 자원을 순수 인적 또는 물질적인것으로만 보아서는 안된다는것이다.

오늘날 자원에 대한 정의에서는 지식경제시대의 성격을 반영하여 인적 또는 물질적원천과 함께 정신적원천도 자원으로 보아야 할것이다.

과학기술이 급속히 발전하고 그에 기초하여 생산이 빠른 속도로 장성하고있는 지식경제시대에는 로력자원, 자연자원 등과 함께 경제정보, 과학기술정보 등의 정보자원도 생산자원으로 등장하여 자원의 중요한 구성부분으로 되고있다.

모든 자원들은 자체의 고유한 형태와 구조 및 특성을 가지고 합법칙적으로 변화발전하며 고유한 운동형태를 띠고있다. 인적자원과 물질적자원, 정보적자원 등은 부단히 형성되고 축적되며 변화된다.

정보자원이란 사람들의 생산활동과 생활에 리용될수 있는 정보원천을 말한다. 정보원천은 자연발생적으로가 아니라 사회적으로 형성리용된다.

정보원천은 정보적의의를 가지는 언어로 표현되는 개별적인 자료, 소식, 지식 등을 의미한다.

정보원천으로서 자료, 소식, 지식 등은 서로 밀접한 련관속에서 생산자원으로서의 기능을 가진다.

정보자원은 고유한 특성을 가진다.

그것은 우선 특수한 존재형태를 가진다는것이다.

다른 형태의 생산자원들은 물질적형태자체의 특성을 가진다.

그러나 정보자원은 문자, 음성과 같은 특수한 물질적매체형태로 존재하거나 지어는 물질적인 매체가 없이 사람의 머리속에 존재할수도 있다. 물질적매체형태로 존재하는 정보자

원의 쓸모도 매체자체의 특성에 의해서가 아니라 거기에 담겨져있는 정보의 내용에 의하여 규정된다.

그러므로 정보자원은 여러가지 매체에 담겨져 정보제품으로 된다. 즉 정보자체는 물질생산부문에서의 제품은 아니지만 일정한 물질형태의 매체에 담겨져 제품화되어 생산자원으로 된다.

정보자원의 특성은 또한 생산물생산과 교환에 간접적으로 이바지한다는것이다.

일반생산자원은 직접 생산에 소비되거나 재생산과정에 참가하여 자기의 가치를 이전시킨다. 그러나 정보자원은 일반적으로 최종적인 제품으로 만들어져 직접 소비할수 없다. 정보자원이 제품생산에 간접적으로 이바지하게 되는것은 그것이 일종의 지식형태의 자원이기때문이다.

정보자원의 특성은 또한 시효성과 기밀성을 가진다는것이다.

일반자원은 수명이 있고 시효가 있다. 정보자원도 수명이 있고 시효가 있지만 정보자원의 사용자체는 그것이 제공되는 시간과 반비례관계를 가진다. 다시말하여 정보자원은 발생한 다음 시간이 짧을수록 쓸모가 보다 커진다. 정보자원은 시효성을 가지는것과 함께 기밀성도 가진다.

정보자원이 시효성과 기밀성을 가지는것은 첫째로, 정보가 해당 시기 각종 사물과 사회적운동의 상태와 그 속성에 대한 객관적반영이기때문이다.

정보의 쓸모는 정보의 내용에 따라 결정된다. 그런데 정보가 반영하는 현실은 끊임없이 변화된다. 따라서 정보가 사장되면 그것이 반영하는 현실과 일정한 차이가 생기게 되며 따라서 그 쓸모는 점차적으로 감소된다.

정보자원이 시효성과 기밀성을 가지는것은 둘째로, 정보분야에서 경쟁이 매우 치열하기때문이다.

국제시장들에서 정보자원이 물질적부의 생산에서 노는 역할로 하여 정보자원의 신속한 획득은 생산자들의 리득에 큰 영향을 미친다. 정보자원의 신속한 획득은 경쟁의 침투화를 가져오며 기밀보장에 최대의 관심을 돌리게 한다.

정보자원이 상품으로 되는 경우 시기나 환경을 놓치면 가치를 상실한다. 따라서 정보자원의 관리에서는 그 사장과 분실이 허용될수 없다.

정보자원의 특성은 또한 여러 사람이 공동으로 리용할수 있다는것이다.

일반자원들은 일단 다른 사람에게 리용되면 그 소유권도 이전되어 제3자가 같은 자원을 리용할수 없다.

그러나 정보자원은 그것이 리용되어도 그 소유권이 반드시 이전되는것이 아니며 반복하여 다른 사람에게도 리용될수 있다. 같은 정보자원이 다른 여러 사람들에게 리용되면 여러 구입자들이 공동으로 리용할수 있다. 그리고 여러차례 반복하여 여러 사람들에게 동시에 리용될수 있다. 같은 정보자원이 여러 사람에 의하여 리용되면 정보가 경제발전에 기여하는 정도는 리용자수에 비례하여 커지며 인류가 공동으로 리용할수 있는 특성을 가진다.

정보자원의 특성은 또한 그 쓸모가 리용과정에 없어지지 않으며 그 리용의 효과성은 가변적이라는것이다.

일반자원들은 그 리용과정에 그 물질적형태가 마모되면서 쓸모도 동시에 없어진다.

정보자원은 그 물질적매체들이 마모될수 있지만 정보의 내용은 복사를 통하여 계속 보

존되고 여러차례에 걸쳐 여러 사람에 의하여 같은 측면에서 혹은 서로 다른 측면에서 반복리용될수 있다. 그리고 정보자원의 쓸모가 그 리용과정에 어느 정도의 경제적효과성을 가져왔는가 하는것은 여러가지 요인들에 의하여 한정할수 없다.

물론 정보자원의 효과에도 한계가 있으므로 시간에 따라 그 효과성이 일정한 정도로 상설될수 있지만 그것은 다른 정보들과 호상 결합되어 새로운 효과를 나타낸다. 정보자원조성에서는 가지고있는 정보가 많을수록 새로운 정보를 생산하고 확대하는 능력도 그만큼 커진다.

정보자원의 특성은 또한 그 쓸모가 잠재적이고 경제적효과성이 불확정적이라는것이다.

일반자원의 쓸모는 그 자연적속성에 의하여 직접 직관적으로 표현된다. 그러나 정보자원의 쓸모는 정보를 담고있는 매체자체의 속성으로 표현될수 없으며 사람들이 정보자원을 리용한 후에야 비로소 객관적효과성을 놓고 평가할수 있다. 그러므로 정보자원의 쓸모는 그것을 사용하기 전까지는 잠정적으로 존재한다.

일반자원리용의 경제적효과성은 쉽게 측정하여 확정할수 있을뿐아니라 그것이 비교적 안정하고 서로 다른 사용자들인 경우에도 큰 차이가 없다.

그러나 정보자원리용의 경제적효과성은 사람과 장소, 시간에 따라 서로 차이를 가질수 있다. 하나의 같은 정보자원도 한 분야에서는 리득이 클수 있으나 다른 분야에서는 그리 크지 않을수 있으며 한 지역에서 효과성은 좋으나 다른 지역에서의 효과성은 떨어질수 있다. 뿐만아니라 오늘 리용하면 큰 리득을 얻을수 있으나 래일 리용하면 그렇지 못할수도 있다.

정보자원의 특성은 또한 다른 자원으로 바꿀수 없다는것이다.

모든 생산자원들은 모두 혹은 일부 다른 자원으로 대신할수 있다. 종이가 나무를 대신하고 수지가 강철을 대신하는것 등은 이미전부터 써오고있다. 이러한 현상은 과학기술이 발전하면서 더 늘어나게 된다. 그러나 정보자원은 다른 자원으로 대신할수 없다.

지식경제시대인 오늘 정보자원은 경제발전에서 매우 중요한 의의를 가진다.

우선 정보자원은 경제발전을 위한 과학기술적진보를 다그쳐 물질적부의 생산을 급속히 늘일수 있게 한다.

지식경제시대에 와서 인민경제의 정보화는 비상이 빠른 속도로 추진되고있다. 로력과 기계설비, 원료, 자재만 있으면 물질적생산을 원만히 보장할수 있다고 생각하던 시기는 지나갔다.

지식경제시대에 생산요소들을 합리적으로 결합하여 리용하고 생산과 류통을 효율적으로 보장하는데서 정보자원의 역할은 더욱 커지게 된다. 그러므로 정보자원을 적극 조성하고 리용하는것은 물질적부의 생산을 더욱 빨리 장성시키기 위한 중요한 담보로 된다. 이것은 새로운 과학기술이 폭발적으로 개발되고 과학기술과 생산의 일체화실현의 촉진, 나라들사이, 경제부문들사이의 경제적연계가 더욱 밀접해지고있는 조건에 맞게 정보자원을 제때에 장악하고 리용하여 생산과 류통에 합리적으로 보장할것을 요구한다.

또한 정보자원은 생산물의 질을 더욱 높일수 있게 한다.

생산물의 질을 높인다는것은 제품을 쓰기 편리하면서도 맵시있고 든든하게 만든다는것을 의미한다.

정보자원은 생산자들의 기술기능수준을 높이고 생산공정을 현대화하고 완비하며 앞선

생산방법을 받아들일 수 있도록 자료를 제공한다. 뿐만 아니라 생산과정에서 기술규정과 표준조작법을 엄격히 지켜 제품의 질을 높이며 생산문화를 철저히 보장하게 한다. 정보자원의 역할은 날이 갈수록 더욱 커지고 있다.

정보자원에서 특히 언어정보원천은 중요한 자리를 차지한다. 언어정보원천에는 언어생활에 리용되는 해당 언어의 정량적, 규칙적, 기술적인 원천이 속한다.

언어정보원천의 정량적원천이라고 할 때 그것은 해당 언어의 모든 어휘, 토, 문장과 같은 것들의 총체적인 수량을 의미한다. 여기에는 입말과 글말에서 쓰이는 문화어만이 아니라 방언, 외래어도 포함된다. 구체적으로는 어휘의 경우 입말어휘, 글말어휘를 포함하여 품사별 총수량, 성구, 속담의 총수량도 속한다.

언어정보원천의 규칙적원천은 언어의 리용과 관련한 모든 원천을 의미한다. 여기에는 언어내적인 문법규칙에 의한 원천과 언어사용과 관련한 규칙과 규범 등에 의하여 표현된 원천이 속한다.

언어정보원천의 기술적원천은 언어정보처리와 관련한 모든 정보기술적원천을 의미한다. 여기에는 정보처리를 위한 형태론적, 문장론적, 의미론적기술 등 언어정보처리를 위한 모든 기술적원천이 속한다.

다른 자연자원은 지구의 력사와 물리적운동과정에 생겨나지만 정보자원은 인류가 생겨나 사회적관계를 맺고 생활하는 과정에 생겨난 것이다.

인류초기에 생겨난 정보자원은 처음에는 입말자원으로부터 글말자원 등으로 발전하여 왔다. 이러한 모든 정보자원은 다 인간이 오직 사회생활을 하는 과정에 부단히 확대되고 그 리용의 효과성이 높아지며 지역적인 차이성과 경계성을 가지고 변화발전한다.

오늘 지식경제시대에 들어와 정보자원은 급격히 늘어나서 그 정보량이 폭발적이라고 할 정도로 방대해졌으며 정보는 시간의 흐름에 따라 더욱 증대되고 있다. 이러한 정보량의 증대에서 주도적인 자리를 차지하는 것은 역시 경제부문과 과학기술부문이다.

그런데 이러한 정보의 대부분은 자연언어로 이루어지고 있다. 이로부터 매개 나라가 세계적인 공통규격에 맞추어 정보자원을 형성하고 관리한다면 증대되는 정보의 처리와 획득에서 보다 큰 효과를 얻을 수 있을 것이다.

실례로 기계번역을 개발할 때 개발에 필요한 원천언어와 목적언어의 단어나 문법지식과 같은 정보자원들은 정보자원시장을 통하여 구입할 수 있다.

이렇게 세계적인 정보자원시장에서 구입하는 정보자원들은 시간과 로력을 절약할 수 있게 할 뿐만 아니라 그 질도 담보하고 있다.

지식경제시대의 요구를 반영하여 최근시기 자연언어처리에서는 정보자원을 효과적으로, 대량적으로 리용하고 공유할 목적으로 언어자료들을 수집하여 축적하는 작업을 진행하고 있다.

이러한 과정에 나온 것이 코퍼스이다.

오늘 코퍼스는 《전자화된 대량의 언어자료모임》, 《언어분석용자료모임》 등의 다양한 의미를 가지고 언어학분야와 자연언어처리분야에서 널리 리용되고 있다.

실례로 언어구조의 자동분석과정에서는 자연언어의 본문자료들에 일정한 정보를 부가해놓은 언어자료모임들을 다루어야 하는데 단어의 형태구조와 문장의 결합규칙 및 의미론적해석규칙, 언어적단위들의 통계적성질들을 밝혀내기 위하여 본문자료들에 자동처리와 관

런한 정보들을 부가해놓은 자료기지(용례사전)를 작성한다. 바로 이러한 자료기지가 코퍼스로 된다.

코퍼스는 한마디로 《언어학적인 정보들이 부여된 실례형의 대규모언어자료기지》로 정의할수 있다.

코퍼스는 본문자료들을 일정하게 배열해놓고 그에 부가된 정보들을 통하여 일정한 지식을 얻는다는 점에서 사전으로서의 기능도 수행하며 각이한 정보자원으로서 리용된다.

코퍼스에는 사람들의 실지 언어사용을 반영한 소설이나 잡지, 사람들의 대화나 담화 과정에 이루어진 모든 언어적정보들이 다 들어있다.

코퍼스에는 소설이나 잡지, 신문기사, 보고문과 같이 문자로 이루어진 코퍼스와 사람들의 말이나 담화, 랑독 등과 같이 음성으로 이루어지는 코퍼스가 있다.

지식경제시대에 와서 사람들은 실지 언어행위에서 진행된 본문자료와 음성자료에 기초하여 필요한 정보들을 컴퓨터로 순식간에 검색, 추적, 리용할뿐아니라 그것을 자연언어 처리를 비롯한 여러가지 목적에 리용할수 있게 되었다.

특히 컴퓨터의 연산속도나 기억용량 등의 기술적문제들이 해결되고 방대한 언어자료를 가공추적할수 있게 된 1990년대부터 코퍼스에 대한 관심이 높아지고 그것을 연구분야에 광범히 도입리용하기 시작하였다.

현재 코퍼스에 대한 연구는 세계적범위에서 진행되고있으며 코퍼스라는 새로운 분야가 개척되었다.

오늘날 언어는 이처럼 정보자원으로 등장하여 경제발전을 추동하는 수단으로 끊임없이 발전하고있다.

우리는 앞으로도 정보자원과 코퍼스에 대한 연구를 더욱 심화시켜야 할것이다.